HEALTH EFFECTS INSTITUTE

Includes Errata Sheet Of 11-01-01

Part 1: Replication and Validation

Daniel Krewski, Richard T Burnett, Mark S Goldberg, Kristin Hoover, Jack Siemiatycki, Michal Abrahamowicz, Warren H White, and Others

∞ *Errata* ∞

Created 11/01/01

Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality

A Special Report of the Institute's Particle Epidemiology Reanalysis Project

Final version, July 2000
Posted to the HEI Website 10/27/00

Page 161. Part II. Caption for Figure 5 should read:
City-specific relative risks in the ACS Study.

Page 162. Part II. Caption for Figure 6 should read:
Shape of concentration-response function (with standardized residuals plotted) for cities in the ACS Study.

Page 174. Part II. Table 32. After "O_3 (ppb)" in the left column, append footnote [b] that reads:
"[b] Based on daily 1-hour maximum concentrations."

Page 178. Part II. Table 33. For O_3 (second row from bottom), in the column "Description of Covariate and Source of Data", the entry should read exactly like the other three:
"Daily average concentrations averaged by year for 1980; from residential, commercial, or mobile monitors"

Page 259. Health Review Committee's Commentary. ***Gaseous Copollutants*** section. The third sentence should read:
"For four gaseous copollutants (carbon monoxide, nitrogen dioxide, ozone, and sulfur dioxide), city-specific annual means of daily average concentrations from the year 1980 were obtained from AIRS and used in the reanalysis (see Appendix E, Part II)."

At the end of the same paragraph, add this sentence:
"For this analysis, the ozone values were based on daily 1-hour maximum concentrations."

Part II, Appendix E (available on request)
Page 5. ***Gaseous Copollutants*** section. The second sentence should read:
"Daily average concentrations of NO_2, sulfur dioxide, ozone, and carbon monoxide were obtained from 1980 to 1989, in addition to the daily one-hour maximum concentrations of ozone."

Part I: Replication and Validation

Daniel Krewski, Richard T Burnett, Mark S Goldberg, Kristin Hoover,
Jack Siemiatycki, Michal Abrahamowicz, Warren H White, and Others

THE HARVARD SIX CITIES STUDY

The Harvard Six Cities Study (hereafter referred to as the Six Cities Study) is a unique, long-term, prospective cohort study designed to evaluate the health effects of exposure to various airborne pollutants. The present reanalysis focused only on that portion of the entire Six Cities Study in which the Original Investigators analyzed an epidemiologic association between mortality and air pollution levels measured from 1977 through 1985, the results of which were reported in the *New England Journal of Medicine* (NEJM)* by Dockery and associates (1993)†. For that epidemiologic analysis, the study population consisted of a random sample of 8,111 white men and women who were between the ages of 25 and 74 years and who resided in one of six US cities at the time of enrollment: Steubenville OH, St Louis MO, Portage WI, Topeka KS, Watertown MA, and Kingston-Harriman TN (hereafter referred to as Harriman).

The data used in the Six Cities Study were derived from questionnaires completed by participants at their time of entry into the study, starting in 1974. Data were also obtained from follow-up questionnaires completed 3, 6, and 12 years after the time of enrollment. The questionnaires were used to elicit information about age, sex, weight, height, education level, smoking history, occupational exposure, and medical history (examples of original and follow-up questionnaires and the coding guidelines are included as Appendix C).

Mortality was assessed during 14 to 16 years of follow-up (totaling 111,076 person-years of follow-up) and 1,430 deaths among the 8,111 subjects were ascertained. Mortality status was determined using information collected from mailings to subjects and by searching the National Death Index (NDI) for the period 1979 through 1989. Underlying causes of death were coded according to the *International Classification of Diseases, Ninth Revision* (ICD-9) (World Health Organization 1975). Deaths from respiratory diseases (ICD-9 codes 485–495), cardiovascular diseases (ICD-9 codes 400–440), lung cancer (ICD-9 code 162), and deaths from all other causes were analyzed separately. These causes of death were coded by an external, certified nosologist not affiliated with the research team. The development of an air pollution database formed an integral component of the original study. Within each of the six communities, ambient concentrations of fine particles ($PM_{2.5}$), total suspended particles (TSP), sulfur dioxide (SO_2), ozone (O_3), nitrogen dioxide (NO_2), and sulfate (SO_4^{2-}) were measured at a centrally located air monitoring station established specifically for the Six Cities Study. Long-term mean concentrations for each pollutant were calculated for periods that were consistent among the six cities. Concentrations of fine particles were collected from 1979 through 1985.

Survival analysis was used to evaluate the association between air pollution and mortality. Life-table survival probabilities for each year of follow-up were estimated for each city, and differences between city-specific mortality rates were assessed using the log-rank test. Cox proportional-hazards models were used to estimate mortality rate ratios for airborne pollutants while simultaneously adjusting for potentially confounding variables. These variables included cigarette smoking, level of education, body mass index (BMI), and occupational exposures to gas, fumes, or dust. In these models, the subjects were stratified according to sex and 5-year age groups, thereby

* A list of abbreviations and other terms appears at the end of the Investigators' Report.

† The original article appears in its entirety at the end of this Special Report.

This is one section of an HEI Special Report that includes an HEI Statement about the research project, a Preface to the Particle Epidemiology Reanalysis Project, the Investigators' Report (Introduction, Summary, Part I, and Part II), a Commentary by the Institute's Health Review Committee, and the Original Articles and Comments on the Reanalysis from the Original Investigators. Correspondence concerning *Part I: Replication and Validation* may be addressed to Dr Daniel Krewski, Professor of Epidemiology & Statistics, Department of Epidemiology & Community Medicine, Room 3229C, 451 Smyth Road, University of Ottawa, Ottawa Ontario K1H 8M5, Canada.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R824835 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

permitting the specification of a baseline hazard within each stratum of sex and age.

AUDIT OF STUDY POPULATION DATA

Data Provided and Source Documents Accessible for the Data Audit

Many of the personnel who were key to the Six Cities Study were still available at Harvard School of Public Health (HSPH) at the time of this reanalysis. Dr Douglas Dockery and Ms Martha Fay (among others) were available to answer questions and to locate relevant data and records. In planning for the data audit and throughout the site visits, the Audit Team (see Appendix A) had the full and generous cooperation and assistance of the HSPH staff.

The original Six Cities Study protocol was not found in the archives and could not be supplied by the Original Investigators. Nevertheless, the Original Investigators provided the Audit Team with a Statistical Application Software (SAS) electronic data file (referred to herein as "Mort6C.file"), which was a copy of the Six Cities database (referred to herein as "Mort6C/HSPH.file") that had been used for the mortality and air pollution analyses. The Original Investigators also supplied a copy of the code book describing each of these variables. (At least three different formal code books had been used during the Six Cities Study.) The Mort6C.file did not contain any information that could be used to identify the individual study participants.

Records were provided during site visits that contained individual identifier information. These included completed questionnaires, subject tracking sheets (known as "pink sheets" for their color), follow-up postcards, death certificates, spirometry sheets, and printouts of computer programs. These records included names, addresses, Social Security Numbers (SSNs), lifestyle habits, and medical history with spirometry printouts, cause of death, names and addresses of relatives, and place of burial. The Audit Team was able to link these records while on site to the Mort6C.file, which did not contain individual identifier information. The Original Investigators provided study participants with several written assurances that confidentiality of these records would be maintained throughout the study. Therefore, the Audit Team agreed to be bound by these same confidentiality requirements. No original records, copies, or notes pertaining to individual identifiers were removed from the site of the audit. Even subject identification numbers (SIDs) were considered confidential and no reference was made to these records in any audit reports.

Existing quality assurance (QA) audits that had been carried out during the course of the study also were made available to the Audit Team.

Sampling the Dataset and Assessing Error Rates in the Original Data

Subjects had been selected in each of the six cities at random using household voting lists, private census lists, partial blocks from street lists, or alphabetized name lists. The Audit Team did not audit the methods for the selection of subjects in the study because none of the source documents could be located, and because the methods have been described in great detail by Ferris and colleagues (1979), including the methods used for minimizing biases in selecting subjects in each city (see Table 1 in Ferris et al 1979).

The Audit Team conducted data audits using two subsets of 250 subjects, each randomly selected. Some subjects happened to be randomly assigned to both audit samples: the subset of the study population and the subset of deceased subjects. This provided some overlap between the two subsets, which functioned as a check on the auditing system.

We chose this sample size for three reasons:

- it would ensure virtual certainty of finding some errors even if the true error rate was as small as 1%;
- it would be sufficiently large to distinguish between error rates of 1% and 5% with reasonable confidence; and
- it would produce quite accurate estimates of error rates, usually within two to four percentage points of the true value.

Original Investigators' Internal Procedures

Questionnaires and mortality records had been thoroughly audited by Ms Fay and internal reports dated February 11, 1981, and March 2, 1981, were made available to the Audit Team. These reports described the scope of the internal audits and the problems found in the study on a variable-by-variable basis. At the time of the first internal audit, error rates by variable ranged from 0% to 23.6%, largely due to inconsistent coding. After corrective actions were taken, the second internal audit showed that the error rates generally fell in the range of 0% to 1% for the majority of variables. These audit reports described the nature of the errors and the decisions made about corrective actions. Some errors noted were so minor in nature that they would not be expected to affect the integrity of the study or the results. In some cases, the documentation showed that decisions were made not to correct variables

for which the error rates were low in frequency. It is clear from these internal audits that most errors were functions of the evolution of the forms used in the study. For example, Ms Fay had found that the education variable on Form 1-71 had an error rate of 18.6% due to a reformatting problem in the fine gradations of some educational levels. There was an inconsistency between the forms as to whether sixth grade constituted the end of grade school or the beginning of high school. The Original Investigators considered the possibility of reformatting the original database, but decided not to because these fine gradations were not relevant to the statistical analyses to be conducted in the future.

These audits demonstrated to the Audit Team that during the conduct of the study, the investigators were concerned with issues of data quality and that they took the steps necessary to eliminate or reduce the impact of these problems.

Original Investigators' Data Collection and Computer Processing

The Audit Team evaluated the documentation of data collection procedures while auditing the questionnaires (administered at baseline) and death certificates, and verified for each subject in the two audit subsets the recorded value of each variable.

For the questionnaires and mortality data, coding conventions and rules were generally quite clear and well documented. As the forms in the study changed, the methods for interpreting the data using established coding conventions and rules were also clear. The resolutions of any discrepancies in coding were well documented. "Missing" data points were handled consistently. In the beginning of the Six Cities Study (late 1970s and early 1980s), data were recorded via handwritten records, typed documents, and computer punch cards; in later years, many versions of computer software were used to record information and data. For the questionnaires and mortality records, the Audit Team was able to start with questionnaires or death certificates and follow the data trail to the Mort6C.file.

Subset of Study Population: Questionnaires

Different versions of the questionnaires were used in different years and different locations in the study. For Watertown, Harriman, and St Louis, the earliest questionnaire was Form 1-71. For Steubenville and some subjects in Topeka, the earliest version was Form 77 (1-76). For the remaining subjects in Topeka and all of Portage, Form 77 (1-76) or Form 78 (1/77) was used. [Form 78 (1/77) and follow-up Form 82 (8/81) are included in Appendix C.] Revisions appeared to have been made to facilitate the accurate recording and coding of responses. Early forms allowed for ambiguous responses, particularly in the occupational exposure sections.

Table 1. List of Questionnaire Variables for Reanalysis Team to Audit and the Criteria for Declaring Errors in the Six Cities Study

Original Questionnaire Variable	Subvariable	Criteria[a]
Subject identification number	Match with city and questionnaire (also match with other records)	Any difference
Sex		Any difference
Exposure to dusts	Total years of occupational exposure to dust	Any difference
Exposure to fumes	Total years of occupational exposure to fumes or gases	Any difference
Education	Category assignment (more or less than high school)	Any difference
Diabetes		Any difference
High blood pressure		Any difference
Smoking status	Current-, former-, or never-smoker	Any difference
Current-smoker pack-years		Any difference
Former-smoker pack-years		Any difference
Height		Any difference
Weight		Any difference
Body mass index	Calculated variable that was not audited directly	Same to whole number
Initiation date of subject on study		Any difference
Time-on-study		Any difference

[a] Any difference between the Mort6C.file and the questionnaires.

The Audit Team coordinator met with the Reanalysis Team to determine which variables in the Mort6C.file would be audited against the Six Cities questionnaires. Table 1 presents the list of 15 variables selected for the audit. We also tried to determine criteria for what would constitute "an error" in the original data, but found that an a priori definition was of limited value. We therefore decided to record any difference found between the Mort6C.file and the questionnaires.

For each of the 15 variables chosen for the data audit, we compared the data in the Mort6C.file with the data on the initial questionnaires to verify that the information recorded on the questionnaires had been correctly entered into the database.

Questionnaire Variables The Audit Team reviewed only the data derived from the questionnaires that were administered at enrollment. We could not audit variables for 1 (0.4%) of the 250 study participants because the initial questionnaire for that individual was missing from the file. A check of files directly before and after this folder failed to locate the missing questionnaire. We did find subsequent questionnaires and other documentation for this subject.

Depending on the variable under examination, more than one auditor evaluated each of the remaining 249 questionnaires in the study population subset. In cases of apparent discrepancies between the Mort6C.file and the questionnaire for any variable, we followed a number of steps to verify that a difference actually existed. If the discrepancy could not be resolved in this way, we gave a detailed written description of the discrepancy to study personnel, who consulted computer programs, other documents, or individuals and then provided a response to the Audit Team.

Table 2 summarizes the percentage of errors the Audit Team found in the variables we examined for the questionnaires.

Subject Identification Number We matched each SID from the Mort6C.file with the SID on each questionnaire. Furthermore, we matched SIDs and personal identification on questionnaires to any other records filed for the same subject: other records included postcards, pink cover tracking sheets, and death certificates. The SID contained a code for the city so the SID checking process also confirmed that the individual was assigned to the correct city. We noted no errors in SIDs in any part of the study.

Race We did not formally audit the race of the subjects because "white" was noted in the inclusion criteria and demographic distribution for the study. However, as we reviewed the questionnaires, we noted no instances that did not meet the established criteria.

Table 2. Audit Results for a Subset of the Six Cities Study Population

Variable	Number of Records	Number of Inconsistencies	Percentage
Date of birth	250	0	0.0
Sex	250	0	0.0
Occupational exposure			
Job exposure to dust	249	14	5.6
Total years of exposure to dust	249	0	0.0
Job exposure to fumes or gases	249	15	6.0
Total years of exposure to fumes or gases	249	0	0.0
Education level	250	0	0.0
Diabetes	250	0	0.0
High blood pressure	250	0	0.0
Smoking status	250	0	0.0
Current cigarette smoker (pack-years)	250	0	0.0
Former cigarette smoker (pack-years)	250	0	0.0
Height (meters)	250	8	3.2
Weight (pounds)	250	2	0.8
Body mass index	250	0	0.0
First year of follow-up	250	0	0.0
Last year of follow-up	250	0	0.0
Time-on-study (years)	250	0	0.0

Sex The sex of the subject from the questionnaire was converted to a binary code in the Mort6C.file. We checked each code against the questionnaire. In addition, because we had access to personal identification information and subjects' medical histories, we were also able to informally verify that the coded information in the Mort6C.file was correct. For example, a subject reported to be female might have corresponding sex-specific medical information; also, many names are culturally considered to refer primarily to one gender. Although these were not absolutes (eg, some men have breast cancer, and some women are named "Billie"), they were flags to the auditors to check further into study data to confirm the questionnaire information. We found no errors in this variable in the audit subset.

Exposure to Dust, Fumes, and Gases Information regarding lifetime occupational exposures to dust, fumes, and gases was requested in the section on residential and occupational history, page 2 of Form 1-71 (used for Watertown, Harriman, and St Louis). Industry, job, and materials handled were requested with approximate dates. Information was coded by years of exposure to dust and years of exposure to fumes, and a dichotomous variable was created. During an earlier internal audit of 89 Form 1-71 questionnaires, the investigators had found inconsistencies in the coding of these exposure data (a 15.7% coding error rate for occupational exposure to dust, and a 12.4% error rate for occupational exposure to fumes and gases. The dichotomous variable was not subjected to an internal audit).

We audited the data for occupational exposure variables against information listed on the initial questionnaire. We found the highest percentage of inconsistencies in the coding of occupational exposure to dust, fumes, and gases. Most of the coding errors in these variables were from the earliest form of the questionnaire, used in Watertown, Harriman, and St Louis. The section for occupational history on Form 1-71 allowed for variability in the way the interviewer recorded information. We found start and stop dates for exposure difficult to determine because no space had been provided on the form for the interviewer to summarize years of exposure. Of the 14 coding errors for the dust category, 12 involved the early questionnaire. Of the two errors in the later version, one was a rounding error. For the exposure to fumes category, 13 of 15 coding errors involved the first questionnaire. The two errors we noted in the later version were both due to rounding errors.

On the revised questionnaire used for Steubenville and for some respondents in Topeka [Form 77 (1-76)], a "years of exposure" column was added for dust and fumes and this information had been directly coded. The Audit Team found some inconsistencies in rounding of data. The only frank error we identified was that on one questionnaire, 26 years of exposure to coal dust had not been noted in the summary column and was not captured electronically.

We found exposures to dust in offices, schools, and libraries to have been coded inconsistently. For example, two long-time teachers had been coded as "0" exposure to dust, whereas another had been assigned a dust code representing "40" years, and dust in a library had been coded for another subject. No criteria used to classify exposures were mentioned in the code books.

Other potential inconsistencies included a bookkeeper in a service station with a code for "7" years of exposure to fumes (carbon monoxide). Another subject's 5 years of employment as a service station attendant had not been coded (this subject had other exposures to fumes for 8 years). In most cases, the subject's description of "materials handled" guided the coding, even if the information was not consistent with the job title. For example, a long-time carpenter did not mention dust exposure and had been coded as "0". A construction worker did not mention exposures and had also been coded as "0" for dust. The Audit Team did not note these as errors because the code book guidelines were to code information in the "materials handled" column. Nevertheless, we noted that the "0" codes for occupational exposures were not necessarily accurate descriptors.

The Original Investigators collapsed the fumes and gases exposure data into a binary variable of yes-no occupational exposure. Therefore, the rounding errors and questions about duration of exposure would not have affected this binary variable. However, recording a "0" for occupational exposure in cases such as the carpenter and construction worker would have influenced the binary categorization. The Audit Team questioned the assignment of "0" for dust exposure in 7 cases and for fumes and gases in 9 cases.

In summary, the Audit Team found (1) 14/249 (5.6%) inconsistencies for occupational exposure to dust: 1 rounding error, 2 overestimates of exposure, and 11 underestimates of exposure; and (2) 15/249 (6.0%) inconsistencies for occupational exposure to fumes: 2 rounding errors, 2 overestimates of exposure, and 11 underestimates of exposure. Underestimates typically had been coded "0" years.

Education As previously discussed, contemporary internal audits showed errors in recording levels of education because of the forms used and the fine distinctions present on the questionnaires. Form 77 (1-76) (used for all of Steubenville and some subjects in Topeka) contained a misprint so that code "1" meant "grade school not completed". The older Form 1-71 (for Watertown, Harriman, and St Louis) used code "1" to mean "grade school completed". Some interviewers using Form 77 (1-76) had crossed out the word "not" and coded this as "1" for "grade school completed" to make it consistent with the previous form. The Audit Team found several instances of this.

Diabetes Subjects were asked if their doctor had ever said they had diabetes or if they had been told they had sugar in their urine. No errors were found in the 249 questionnaires examined.

High Blood Pressure Subjects were queried as to whether they had been told their blood pressure was high and if

they had been treated for it in the last 10 years. In one case, the auditors concluded that notes written on the margin of the questionnaire suggested that a woman who had been coded as not having high blood pressure was likely treated for hypertension. This was not considered a coding error.

Smoking Status Subjects in this study were classified as current-smokers, former-smokers, or those who never smoked. This variable referred only to cigarettes because the coding protocol allowed cigar and pipe smokers to be classified as nonsmokers. We checked the Mort6C.file for each of the 249 subjects to determine that subjects classified as "nonsmokers" had no history of cigarette smoking, that "former-smokers" had matching data for former-smokers, and that "current-smokers" were matched with current smoking data. We found no differences in this variable.

Pack-Years for Current-Smokers and Former-Smokers An internal audit from 1981 showed that the calculation of pack-years of smoking cigarettes had been somewhat inconsistent in this study. The rules for calculating this variable had not been followed closely, especially for data from the earliest Form 1-71 with regard to "total amount of cigarettes currently smoked" and to "periods of smoking abstinence". Early calculations appear to have introduced a six-month correction factor to address the idea that people probably did not begin smoking on January 1 of a given year and did not stop smoking on December 31. Smoking data for respondents who initially completed Forms 77 (1-76) and 78 (1/77) were different from those for subjects who were interviewed using Form 1-71 because the six-month correction factor was dropped from later calculations. Furthermore, this study included a number of subjects who smoked their own hand-rolled cigarettes, and the use of hand-rolled cigarettes was factored into the total consumption. The 1981 internal audit clearly described limitations in how these problems in smoking data could be addressed. It concluded that the change from Form 1-71 to Forms 77 (1-76) and 78 (1/77) resulted in an underestimate of smoking pack-years by about 3% in the three cities where Form 1-71 was used (Watertown, Harriman, and St Louis).

The Audit Team spent a considerable amount of time resolving issues about smoking data. We discussed with Dr Dockery and Ms Fay the rules and formulas for recalculating pack-years, and then performed recalculations on the basis of these discussions and the documentation present in the code books. The Audit Team confirmed the findings of the 1981 internal audit; specifically, a slight underestimate of smoking for former-smokers versus current-smokers due to changes in the forms, and a slight underestimate (approximately 3%) of pack-years of smoking in the study.

Height, Weight, and Body Mass Index Height and weight were measured by the interviewers and recorded manually on the questionnaires. We audited height and weight against the Mort6C/HSPH.file because the printout of Mort6C.file provided to the Audit Team supplied only the aggregate calculation of BMI. After the audit, values for height and weight from the Mort6C/HSPH.file were validated against the Mort6C.file.

The audit of the height variable revealed six instances in which the Mort6C.file and the questionnaires differed. One was a simple rounding error; in the other five cases, the data file had been changed because subsequent spirometry measurements or questionnaires showed that the initial measurements of height had been inaccurate.

The audit of the weight variable revealed two differences, of which one was a simple rounding error. The other was for one of the subjects whose initial height measurement had been recorded incorrectly and changed later. Likewise, the subject's weight had been changed from 121 to 140 pounds. During the data editing phase, corrections were made to the data by the investigators whenever possible. It is possible that this change in weight was made during the editing process. Given the changes in data for this subject, we concluded that the original height and weight data had accidently been recorded in opposite fields for this individual.

These differences demonstrate the Original Investigators' attention to the consistency of data over time and have no negative impact on the study's results. Our recalculation of BMI revealed that differences were due only to the height and weight values as discussed. Our recalculation of the overall mean BMI for each city, as reported in Table 1 of the NEJM publication (see Table 17a), showed very minor differences.

Initiation Date of Subject on Study We crosschecked the date of enrollment into the study against the date of the interview on the initial questionnaire, the Mort6C.file provided to the Reanalysis Team, and the precursor file at HSPH (Mort6C/HSPH.file). For one subject, the month reported on the questionnaire was poorly legible; it appeared to us that the handwritten date of the interview could be November instead of December. The December date appeared in the Mort6C.file and in the Mort6C/HSPH.file. All other enrollment dates matched in their entirety (mm/dd/yr) for the audit subset.

Time-on-Study (Initiation Date and Last Date) We could not audit "time-on-study" directly because it was a calculated variable. The calculation depended upon what cutoff date had been used for each city. Another factor in verifying these calculations was that some records had been updated after the ending date for the study analyses.

If the Audit Team found that the subject had died between the dates of completing the initial questionnaire and the last date of follow-up for that city, we verified the date of death against the death certificate (or, in cases where no death certificate was available, against information supplied by the subject's family) and calculated the time-on-study accordingly. To audit this vital status variable for subjects who had not died during follow-up, we used information on dates from the last completed questionnaire, the pink cover tracking sheet, work cards, some summary computer printouts, and postcards that were sent periodically to study participants and returned by them. We compared all of this information against interim printouts from the Mort6C/HSPH.file.

After we completed the audit for time-on-study in Watertown, several discrepancies were noted in the data for each of the other five cities. When we discussed this with Ms Fay and Dr Dockery, a search of their records showed that an error in a computer program had resulted in some data for some subjects not being updated in each of the other five cities. This led to a loss in the total number of years of follow-up. (In epidemiologic studies, this is referred to as "early censorship of person-years of follow-up".) For the Audit Team's subset of 249 subjects, Dr Dockery and Ms Fay re-created the time-on-study and found a loss of approximately 1% in the reported person-years for the entire study. The Original Investigators also provided a summary of the entire study showing the number of subjects in each city for which early censorship of data had occurred (Table 3). Early censorship was greater in Portage and Topeka than in other cities.

Table 3. Early Censorship of Person-Years of Follow-Up in the Six Cities Study

City	Number of Subjects	Number of Subjects with Early Censorship	Percentage
Harriman	1,258	35	2.8
Portage	1,631	185	11.3
Steubenville	1,351	51	3.8
St Louis	1,296	36	2.8
Topeka	1,239	152	12.3
Watertown	1,336	0	0
Total	8,111	459	5.7

Subset of Deceased Subjects: Death Certificates

The Audit Team randomly selected another independent subset of 250 SID numbers that had been coded as deceased in the Mort6C.file. We examined the 248 (92.2%) of the matching death certificates that were found. We compared the following information in the Mort6C.file and the source documents:

- date of death in Mort6C.file against the date of death on the retrieved death certificate;
- identifying information of subjects contained on the death certificate against the same information on the subject's initial questionnaire so as to determine that the correct death certificate had been obtained for the person who completed the study questionnaire;
- cause of death recorded in the Mort6C.file against the ICD-9 code the study nosologist wrote on the pink cover tracking sheet attached to the death certificate;
- cause-of-death code assigned by the study nosologist against the ICD-9 code interpreted by the Audit Team from the death certificate;
- cause-of-death groupings recorded in Mort6C.file against the criteria for assigning the cause of death to a group;
- date of subject's initiation on study and date of death on the death certificate against calculation of time-on-study.

Date of Death When we matched the Mort6C.file with the death certificates, we found errors for two subjects. One error (year of death) had been detected by the Original Investigators after the epidemiologic analysis had been completed, and the current Mort6C/HSPH.file reflected the correct information. The second error (month of death) had not been corrected in the current Mort6C/HSPH.file.

Correct Death Certificate Using information from the questionnaires, the Audit Team verified that the death certificate on file reflected the correct study participant by matching the full name, SSN, birth date, and gender. Social Security Numbers were not recorded on all death certificates, and the Audit Team noted other minor inconsistencies between the death certificates and the questionnaires, which usually involved one digit of the SSN or birth date. However, the 247 available death certificate and questionnaire pairs matched in enough fields to verify that all the death certificates pertained to the correct study participants.

Table 4. Discrepancies Between Cause-of-Death Codes by Study Nosologist and Audit Team for the Six Cities Study

Code by Study Nosologist	Comments	Code by Audit Team's Nosologist	Change of Code Would Have Altered the Category in the Epidemiologic Analysis
Diabetes with ophthalmic manifestations (250.5)		Diabetes with renal manifestations (250.4)	
Malignant neoplasm without specification of site (199.1)	The death certificate reads, "metastatic ADCA [adenocarcinoma] to liver, unknown primary".	Secondary neoplasm to liver (197.7)	
Congenital mitral stenosis (746.5)	The coding of this case appears to have been in question because one notation in red ink lists 394.0, but then a comment is added that the "patient's age affects the coding". The death certificate reads, "rheumatic heart disease (mitral stenosis)"; rheumatic heart disease is an acquired, not congenital, condition. Therefore, the nosologist's first code of 394.0 is consistent with the death certificate.	Mitral valve stenosis (394.0)	
Chronic obstructive pulmonary disease (COPD) (496.0)	The death certificate lists the following causes of death: line a, respiratory failure; line b, COPD; line c, metastatic malignant melanoma.	Malignant melanoma (172.9)	This death would have changed categories from "cardiopulmonary" to "lung cancer".
Chronic ischemic heart disease (414.9)	The death certificate lists the following causes of death: line a, hypotension; line b, massive stroke; line c, congestive heart failure (CHF). The order listed by the physician is questionable because the underlying (primary) cause of death most likely was the massive stroke, although the physician lists CHF on line c.	Cardiovascular aneurysm (CVA; stroke) (436.0) or CHF (428.0)	
Acute myocardial infarction (410.0)	The death certificate lists the following causes of death: line a, acute myocardial failure; line b, atherosclerotic heart disease; and line c, cancer of kidney.	Malignant neoplasm of kidney (189.0)	This death would have changed categories from "cardiopulmonary" to "other".

Cause-of-Death Codes First, the Audit Team compared the primary cause of death listed in the Mort6C.file as a four-digit ICD-9 code against the nosologist's code recorded on the pink cover tracking sheet attached to the death certificate and found that 100% of the codes matched. In three cases, the Mort6C.file included no ICD-9 code because the death certificate had not been coded.

Two areas on the death certificate record the causes of death: Cause of Death Part I and Part II. Part I has three lines. One, two, or three lines may be completed by the physician as follows: line a, immediate cause of death; line b, explanation of the immediate cause (immediate cause due to or a consequence of); and line c, explanation of line b (due to or a consequence of). The final entry in Part I is considered the underlying (or primary) cause of death. Part II is a one-line area for the physician to detail other significant conditions that are not directly related to the underlying cause of death.

Using the ICD-9, Dr Donna Foliart of the Audit Team coded the underlying (primary) cause of death listed on each of the death certificates and the Audit Team compared them with the study nosologist's ICD-9 code (which had been recorded on the pink cover tracking sheet attached to the death certificate). In six cases, Dr Foliart's code did not match the full four digits of the study nosologist's code. In

Table 5. Audit Results for the Subset of Deceased Subjects in the Six Cities Study[a]

Variable Used in Epidemiologic Analysis	Number of Records	Number of Inconsistencies	Percentage
Date of death	248	2	0.8
Death certificate and study participant identifiers	247	0	0
Nosology code	248	6	2
ICD-9 code in Mort6C.file	248	0	0
Cause-of-death group based on nosologist's code	248	0	0
Total	1,239	8	0.6

[a] All source documents were death certificates.

the epidemiologic analysis, the investigators had grouped deaths into four cause-of-death categories: cardiopulmonary, lung cancer, other, and missing. Of the six discrepancies in ICD-9 codes, two would have altered the category used in the original analyses.

Details of the six discrepancies are described in Table 4, which gives Dr Foliart's code, the study nosologist's code, and comments from the Audit Team. The findings from the audit of the subset of deceased subjects are summarized in Table 5.

AUDIT OF AIR QUALITY DATA

Description of Original Air Quality Dataset

The original epidemiologic analysis characterized ambient air quality as long-term mean concentrations of various air pollutants. The following variables were reported for each of the six cities from measurements taken during the indicated years: concentrations of total particles (1977–1985), inhalable and fine particles (1979–1985), sulfate particles (1979–1984), aerosol acidity (H^+)(1985–1988), sulfur dioxide (1977–1985), nitrogen dioxide (1977–1985), and ozone (1977–1985). Measurements of air pollutants were taken using well established methods augmented with newly developed techniques as necessary. The methods used to calculate mean concentrations (eg, as the average of seasonal means, annual means, or individual observations) were not specified.

Further description of the Audit Team's decisions about which air quality data to audit and how to proceed is presented below for different groups of pollutants.

Gases The gases (SO_2, NO_2, and O_3) had been monitored hourly by standard continuous instrumentation and recorded in parts per billion. The measurements had been checked by contemporary external audits (eg, Eaton et al 1982). Selective inspections by our Audit Team of the original data records, operator logs, and field audits for these measurements did not indicate any unusual problems. As a result, we decided not to audit these data or the findings associated with them.

Acidity Aerosol acidity had been measured for about one year in each city. The hydrogen ion concentrations were determined using research-quality methods to analyze 24-hour fine particle samples collected with Harvard impactors (Koutrakis et al 1988). However, measurements were conducted in only two cities at a time, starting with Harriman and St Louis from December 1985 through August 1986 (9 months) and finishing with Topeka and Watertown in August 1988 (10 and 14 months, respectively). Thus, it was impossible to compare acidity for a common time period.

Furthermore, the acidity data were not necessarily linked with particle data in the same city; for example, dichotomous particle sampling at Watertown ended 18 months before the initiation of measurements of acidity. Because intercity comparisons were confounded by uncontrolled interannual variability, and the acidity measurements were disconnected from other particle measurements, we decided not to audit them.

Particles The Original Investigators reported mean concentrations for four classifications of particles in each of the six cities: TSP (particles with aerodynamic diameters as large as 50 μm), inhalable particles, fine particles, and sulfate particles. In the sections that follow, we describe different samplers and methods of arriving at these four groups. All particle measurements were recorded as mass concentrations ($\mu g/m^3$).

Values of mass for TSP (for the years 1977–1985) and sulfate particles (for the years 1979–1984) were determined from 24-hour samples collected by General Metal Works (regulatory standard) high-volume samplers having unrestricted inlets. The sample was first weighed to determine the concentration and then subjected to chemical analysis to determine the concentration of sulfate ions. The methods used were Federal Reference Methods and they had been subjected to contemporary external audits (eg, Eaton et al 1982) of both the sample collection procedures and the laboratory analyses.

Inhalable particle mass was calculated from coarse and fine particle mass, which had been determined from 24-

hour sample pairs collected by Beckman dichotomous samplers. At the time of its introduction, the dichotomous sampler was relatively new and untested and was still undergoing a number of operational difficulties. Furthermore, most researchers had much less experience with it than they had with the older high-volume sampling technology.

Compared with the dichotomous sampler, the high-volume sampler is a "simple" tube with a single filter mounted in the middle; one end of the tube is open to the atmosphere and the other is attached to a powerful vacuum pump, thus allowing the filter to collect particles of all sizes. In contrast, the dichotomous sampler is designed on the complex principle of virtual impaction. In still air, and under the influence of gravity, large particles settle out more rapidly than small particles. In curving or decelerating airflows, and under the influence of centrifugal forces, large particles are correspondingly quicker than small particles to migrate to the outer boundaries and impact on outer surfaces. The inlets of particle samplers are designed to impose contortions on entering airflows sufficient to make nearly all particles above a selected size impact on the surfaces of the inlet. (This is the principle of the size-selective inlets [SSIs] routinely used to remove from the sample air particles greater than 10 or 15 μm in aerodynamic diameter.) The remaining smaller particles are captured on a fine particle filter.

The dichotomous sampler exploits this same aerodynamic separation phenomenon to separate from the same airstream particles both above and below 2.5 μm in diameter. The filter in the primary flow of intake air (the fine particle channel) collects only particles smaller than 2.5 μm. Most of the intake air (typically 90%) is forced to undergo a sharp deceleration (secondary flow) and is focused into a receptacle of dead air (the coarse particle channel). At the bottom of the receptacle is a coarse filter that collects coarse particles, any directly impacted particles, and any fine particles carried by the secondary air flow. The calculation of coarse particle mass concentration includes a correction factor for the fine particles collected in the coarse particle channel.

In the dichotomous samplers used in the Six Cities Study, the fine particle channel collected particles smaller than about 2.5 μm and the measurement was recorded directly as fine particle (FP) mass. The coarse particle channel collected particles between 2.5 μm and 10 or 15 μm in aerodynamic diameter (the upper bound measurement depended on the inlet size used at the time, which is discussed later). These samples were corrected for the inclusion of some fine particles, and the correction resulted in the coarse particle (CP) mass. Then both FP and CP values were added to yield the inhalable particle (IP = FP+CP) mass, which included all particles smaller than 10 or 15 μm in aerodynamic diameter.

In different years, measurements of mass from dichotomous samples were carried out by different organizations in different laboratories (described in detail in the next section) by two fundamentally different methods. The dichotomous sampler analyses also had not been verified by blinded audits of samples, as had the high-volume sample analyses. In addition, the Audit Team found the existing records of dichotomous samples to be more fragmented than those for the high-volume sampler measurements. For these reasons, we decided the dichotomous sampler particle data ought to be the principal focus of our audit.

Original Analysis of Air Pollutants from the Dichotomous Samplers

Over the course of the study, several changes were made in operating the samplers and in the methods used to analyze the samples.

- Until and throughout most of 1981, the filters from the samplers were analyzed by an EPA laboratory in North Carolina. This laboratory determined mass by β-absorption gauge.
- In October and November 1981 (exact dates varied in each city), the analysis of the filters was transferred to HSPH until 1984. The HSPH laboratory used standard gravimetric analysis in which the filters were weighed before and after exposure. (Courtney and colleagues [1982] found no significant bias between the two methods of sample analysis when they applied them to air quality samples [not from the Six Cities Study] collected in North Carolina.)
- In January and February 1984, the analysis of the filters was transferred from HSPH back to the EPA laboratory in North Carolina; the mass was again measured using the same methods as before.
- Also in January and February 1984, the filters on the coarse particle channel were oiled to improve particle adhesion. This action was taken in response to a discovery that substantial and variable particle losses had been occurring in transit and handling (Dzubay and Barbour 1983; Spengler and Thurston 1983). Oiling the filters would have increased the levels of coarse particle mass but would not have affected measurements of fine particle mass.
- In March and April 1984, new inlets were installed that reduced the 50% sampling cutoff for particle size from 15 μm to 10 μm. This action would have resulted

Table 6. Changes in Dichotomous Sampler Configurations and Analysis Methods in the Six Cities Study

Factor Changed	Epoch 1 (1979–1981)	Epoch 2 (1981–1984)	Epoch 3 (1984–1984[a])	Epoch 4 (1984–1988[b])
Inlet size cutoff[c]	15	15	15	10
Coarse filter	Dry	Dry	Oiled	Oiled
Type of analysis	β Gauge	Gravimetric	β Gauge	β Gauge
Analysis laboratory	EPA	HSPH	EPA	EPA

[a] At the longest, this epoch lasted from January through April of 1984.

[b] The data from 1986–1988 were not used in the epidemiologic analysis published in NEJM.

[c] From the coarse particle channel of the size-selective impactor.

in lower levels of coarse particle mass but would not have affected measurements of fine particle mass.

The Audit Team used these transitions to partition the dichotomous sampler measurements into four distinct epochs, as summarized in Table 6.

Data Transmission, Electronic Recording, and Contemporary Quality Assurance

Quality assurance of data gathering procedures was centrally coordinated at HSPH. As they were being applied in 1982, QA procedures were described in a paper presented at the annual meeting of the Air Pollution Control Association (Briggs et al 1982). A contemporary QA manual (Harvard School of Public Health Air Quality Group 1982) was also available. Both of these documents had been written before any of the changes had been instituted in how the dichotomous samplers were operated and how the samples were analyzed.

From 1979 through the summer of 1981 (Briggs et al 1982), filters from the six cities were returned to the EPA laboratory for analysis. These shipments were accompanied by standard forms (EPA 3B) that supplied information (such as total flow rate and the duration of the sample run) needed to convert the filter loadings to ambient concentrations. The EPA laboratory performed the analysis and the calculations of concentration and returned concentrations corrected for blank filter values. Meanwhile, HSPH collected weekly field logs and calibration records directly from the sampler operators.

The EPA data were screened for encoding and transmission errors, compliance with standard operating procedures and criteria, and statistical anomalies (outliers), and then merged with other study records at HSPH into a master data file. Briggs and colleagues (1982) outlined a review process that augmented each record with diagnostic variables (referred to as "flags") that indicated whether procedures and data were within acceptable ranges.

In the summer of 1982, the Quality Assurance Division of the EPA's Environmental Monitoring Systems Laboratory organized and coordinated a thorough systems audit carried out through personnel of the Research Triangle Institute (Eaton et al 1982).

No updated documentation was located for the years after 1982. The Audit Team assumed that the same procedures were used but likely were modified when gravimetric measurements were made at HSPH.

In addition to the 1982 systems audit described above, the Office of Scientific Integrity (OSI) conducted an external and independent review in response to an internal accusation of misconduct in the processing of ozone measurements. The OSI scrutinized the gas concentration data in detail and concluded that their "exhaustive inquiry resulted in a 'clean bill of health' for the study and for the Six Cities scientists" (SW Hadley, written communication, November 1990).

Data Provided and Source Documents Accessible for the Reanalysis

The Audit Team expected to have available a master electronic database of all air pollution measurements for the entire Six Cities Study; however, master data files were not found. Instead, various data files contained different data subsets, which appeared to have been selected from a common database according to different screening criteria. The efforts to reconstruct the data used to produce the results published in the NEJM are discussed in the next sections.

The primary data that seemed to be missing from the master database were the dichotomous sampler data. Several records documenting original dichotomous sampler measurements and analyses were accessible for some time periods from some cities: laboratory (both EPA and HSPH) transmittals of filter sample measurements and concentration calculations (both electronic and hard copies), some

HSPH data files, and field logs from the dichotomous sampler operators. However, there was no city or time period for which all of these records could be located.

Audit Objectives for Data from Dichotomous Samplers

The Audit Team arrived at the following decisions regarding the scope of the audit for the air quality data:

- We decided not to conduct an audit of the gases because data on gases (SO_2, NO_2, and O_3) had been appropriately checked by external contemporary audits and the 1990 OSI investigation; therefore, no further review was warranted.
- Data on aerosol acidity had not been collected over a common time period for all six cities, and the data had not been necessarily connected to concurrent particle measurements; therefore, an audit was not required.
- The particle data from high-volume samplers had been collected and analyzed with Federal Reference Methods and subjected to contemporary external audits; no further review seemed necessary.
- The particle datasets from dichotomous samplers had been acquired and analyzed with different methods and procedures at different times; these warranted the primary attention and resources of the Audit Team.

Our audit of the air quality data had three broad objectives:

1. verify the conversion of primary filter measurements of air pollutants into concentrations;
2. evaluate the procedures for validating and archiving the concentrations; and
3. clarify how the published means had been derived and evaluate how sensitive the means may be to computational procedures and data selection criteria.

Dr Warren White of the Audit Team conducted two site visits at HSPH on March 8 through 12 and April 12 through 16, 1999. Two years earlier, Dr White had acquired from Dr Dockery a computer spreadsheet containing various particle mass concentrations for 1979–1986: (1) TSP data from high-volume samplers; (2) inhalable particle data from high-volume samplers with SSIs (these data had been recorded every sixth day and had not been used in the air pollution analyses for the NEJM article); and (3) fine and coarse particle data from dichotomous samplers. This extracted dataset (referred to hereafter as 1997.file) had been assembled specifically for Dr White and the data included had not necessarily been selected according to the same criteria used for the epidemiologic analysis of mortality and air pollution. In preparing for the Audit Team's site visit, Dr White used this 1997.file to guide which measurement locations (cities) and periods would be appropriate to review in detail at HSPH.

Objective 1. Verify Conversion of Primary Filter Measurements into Concentrations

To convert a simple filter measurement to an ambient mass concentration, one generally needs four numbers: the mass of the filter with the sample, the mass of the (blank) filter without the sample, the sampler flow rate, and the sampling duration. We wanted to recalculate a few filter measurements to establish the following points.

- The correctness of the calculation, which is significantly more complex for a sample from a dichotomous filter than one from a simple filter. The Audit Team also noted that the equation for this calculation had been reported incorrectly in HSPH's QA Manual for Air Quality Assessment (section III, chapter 6, page 4, C_MASS formula, May 1982); therefore, we wanted to verify the actual methods used for these calculations.
- The handling of the blank correction factor and its effect on uncertainty. The QA Manual states (in section I, chapter 10, page 1, March 1982) that the first filter in each tray of 36 was to be used as a blank in the analysis, but it also indicates (in section III, chapter 6, pages 3 and 4) that mass concentrations were to be calculated from β-absorption gauge measurements with no blank corrections.
- The reporting convention for concentrations: ambient conditions, standard temperature and pressure, or something else?

Extant Original Records by Epoch The archival master electronic data files described by Briggs and colleagues (1982) were not found for epoch 1 (1979–1981). Contemporary hard copies were located for at least some of the concentration transmittals received at HSPH from the EPA laboratory during epoch 1 (1979–1981), and the Audit Team was able to review several monthly records from Harriman (1980), Portage (1980), Steubenville (1980), and St Louis (1980). The EPA transmittals describe data before they were subjected to the screening process described by Briggs and colleagues (1982) and are therefore unflagged.

Similarly, master data files for epoch 2 (1981–1984) were not found. The only laboratory records available for inspection for epoch 2 were those from the HSPH laboratory.

Printouts of the master data files were located for epochs 3 and 4 (1984–1988). These printouts had been produced at HSPH in the late 1980s and accounted for essentially all of the observations recorded in the site operator logs the Audit Team reviewed (Harriman 1985; Topeka 1984, 1985,

Table 7. Original Records for Dichotomous Samplers Examined in the Audit of the Six Cities Study[a]

City	1979–1981	1982–1983	1984–1988[b]
Harriman	EPA Lab		Field logs, HSPH MF
Portage	EPA Lab		
Steubenville	EPA Lab		HSPH MF
St Louis	EPA Lab	Field logs, HSPH Lab	HSPH MF
Topeka	Field logs		Field logs, HSPH MF
Watertown			Field logs, HSPH MF

[a] Individual entries in columns represent samples of records spanning several weeks to several months, not all of the years mentioned. Different datasets were available in different years. EPA Lab = EPA laboratory transmittals; HSPH Lab = HSPH weighing laboratory records; HSPH MF = HSPH master data files; and field logs are from dichotomous sampler operators.

[b] The data from 1986–1988 were not used in the epidemiologic analysis published in NEJM.

and 1988; Watertown 1985). The printouts were of data files that had been subjected to the QA procedures described by Briggs and colleagues (1982) and included flagged data fields.

The Audit Team received no response to a request to visit the EPA contract laboratory in North Carolina. Table 7 summarizes the original records the Audit Team examined during the site visits at HSPH. We did not randomly sample cities and periods, as we did with individual health records, because air quality data records were not uniformly available.

Audit Team's Recalculations of Concentrations We could not recalculate the measurements made during epochs 1, 3, and 4 because records of the analyses completed at the EPA laboratory were not available at HSPH. These data conversions should not be of concern, however, because the EPA laboratory was the leading practitioner of these methods at the time.

The Audit Team was successful in recalculating concentrations from primary filter measurements for some of the analyses conducted at HSPH during epoch 2. Figure 1 shows results obtained for 30 observations of concentrations for St Louis from May through July 1983. The Audit Team found no indication that adjustments were made for variations in temperature and pressure. The root-mean-square difference between calculated and reported concentrations is 0.7 μg/m^3 for fine particles and 1.0 μg/m^3

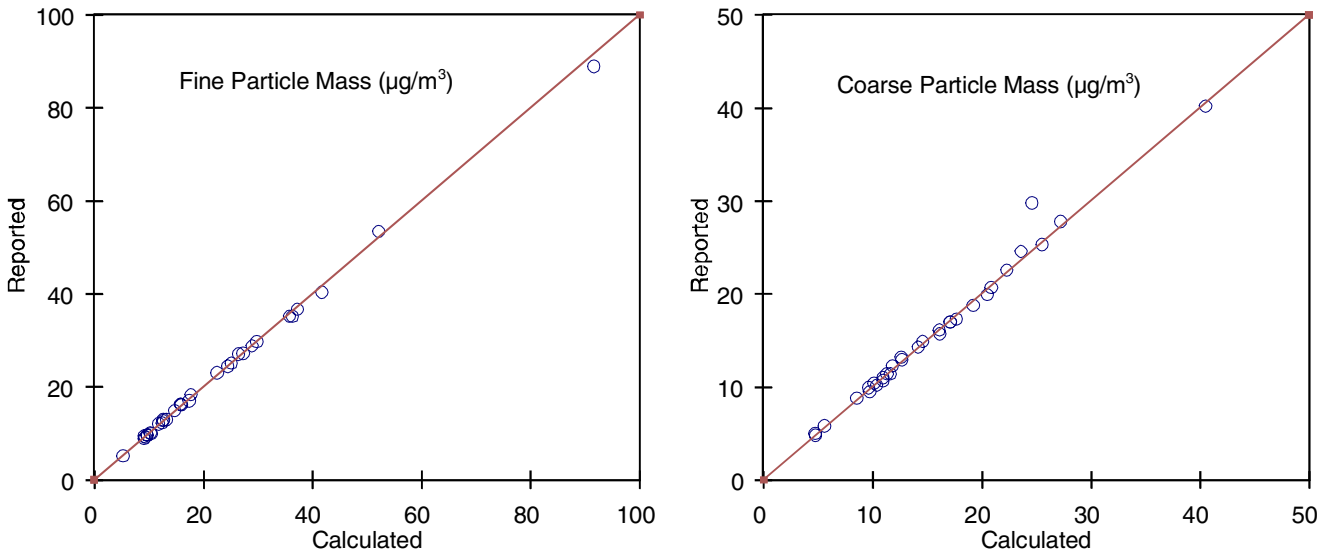

Figure 1. Agreement between reported and newly calculated fine (left panel) and coarse (right panel) particle mass concentrations in St Louis May–July 1983. The straight line in each panel defines perfect agreement.

for coarse particles. This level of discrepancy could arise from minor uncertainties as to the exact procedure used in the original conversion.

Objective 2. Evaluate Procedures for Validating and Archiving Concentration Measurements

The written procedures used to validate and document laboratory transmittals and the computerized and manual review processes used to inspect the data and input them into the master air pollution files are described in the section Data Transmission, Electronic Recording, and Contemporary Quality Assurance. That section also references a contemporary QA manual (Harvard School of Public Health Air Quality Group 1982); the Audit Team found some of the formulas and descriptions in that manual to be clearly erroneous. The manual refers to an additional report documenting procedures at the EPA contract laboratory, but no copy of that report could be located at HSPH.

One set of records the Audit Team examined included hard-copy transmittals from the EPA laboratory of data from Steubenville for the period April 1979 through February 1981. This period had been discussed in some detail by Briggs and colleagues (1982) to illustrate the conventions for validating data. According to Briggs, during the period of September 16 through 24, 1980, the samples for all 9 days had been noted by the site operator as "suspect" because of repairs to the roof on which the sampler was located. Concentrations for all of these samples were included in the EPA transmittal to HSPH, as they should have been. According to Briggs' documentation, the data were not voided but were archived and coded with a "suspect" flag.

Objective 3: Clarify Derivation of Published Means and Evaluate Their Sensitivity to Computational Procedures and Data Selection Criteria

Air Quality Dataset The master air pollution data files were no longer accessible on the HSPH computer system and the staff of HSPH were unable to locate a copy of this file. Various electronic data files examined during the first visit were found to contain different data subsets and appeared to have been selected from a common database according to different screening criteria. However, one source of potential problems was that different values were sometimes reported in different files for the same observation.

During the first site visit, Dr Dockery produced a provisional and incomplete reconstruction of the air quality data used in the NEJM analysis; he supplemented these data with dichotomous sampler mass concentrations used in a time-series analysis published in a 1996 article in the *Journal of the Air and Waste Management Association* (JAWMA; Schwartz et al 1996; hereafter, the electronic file containing the data published in JAWMA is referred to as JAWMA.file). None of the data files found on the HSPH computer and none of the reconstructed databases could produce the exact air pollution concentration averages reported in the NEJM article. Before the second onsite audit, Dr Dockery produced an improved reconstruction of the NEJM analytical file (hereafter referred to as Reconstruct.file), which was the one the Audit Team used to compare with all other original records of air pollution transmittals. The Reconstruct.file likely comprised electronic data files extracted from the master air pollution files in different years, according to criteria that evolved with time.

Comparison of Original Records with Reconstruct.file The Audit Team verified the fine particle mass concentrations in the Reconstruct.file with some of the original records described in Table 7 for each of the epochs described in Table 6; the results are summarized in Table 8. The Audit Team could account for all but 3 of the 1,010 values examined in the Reconstruct.file (in Table 8, see the column "NEJM vs Original Records" under "Number Unmatched"); these 3 data points could simply have been missed in the audit.

Comparison of Original Records with JAWMA.file Although the JAWMA air data were not formally audited, Table 8 includes results of a similar comparison for the JAWMA.file because it is discussed below as an alternative representation of the dichotomous sampler data. A significantly larger number (64 of 1,191) of the JAWMA values that were examined could not be accounted for and some of them are from dates when field logs indicate that no samples were taken.

Criteria for Selecting Data for the Mortality and Air Pollution Analysis No contemporary account could be found of the criteria used to select data for the mortality and air pollution analyses. Nevertheless, the Audit Team was able to infer some of the criteria used by comparing the Reconstruct.file with available earlier records. This comparison clearly reflected that some selection criteria had changed over the years, as described in the next sections.

Restriction on Coarse/Fine Mass Ratio Data from epoch 1 (1979–1981) were systematically excluded whenever the coarse/fine mass ratio was less than 0.3 or greater than 1.3. This restriction reflects early EPA guidance; Briggs and colleagues (1982) noted that it does not allow for actual variations in particle size distributions: thus, it "appears to be an undesirable check for bad values in its present form. … [I]f this criterion were employed to void data, it would likely introduce bias into the datasets." Data from later

years (1982 on) were included regardless of coarse/fine mass ratios in accordance with the recommendation of Briggs and colleagues. The abrupt elimination of the coarse/fine mass ratio restriction is shown in the time-series ratios reported for Portage, which are plotted in Figure 2 from the Reconstruct.file. A similar pattern was found in all six cities.

Even during the time it was applied, however, the coarse/fine mass ratio restriction did not greatly affect the fine particle concentrations for Portage in any obvious manner. This effect is shown in Figure 3, which compares the Reconstruct.file data from HSPH for Portage in 1980 with the values reported by the EPA contract laboratory. The EPA data points that are unmatched by HSPH data points are those values that HSPH excluded because the coarse/fine mass ratio fell outside the applied boundaries.

The Audit Team also assessed the empirical effect of the coarse/fine mass ratio restriction on average concentrations by applying the restriction to otherwise unrestricted data in the Reconstruct.file for 1982 and later years (Table 9). Had the restriction been applied to the data in these years, the greatest impact would have been seen in Topeka, where Briggs and colleagues (1982) reported the average measured ratio would have fallen outside the "appropriate" range.

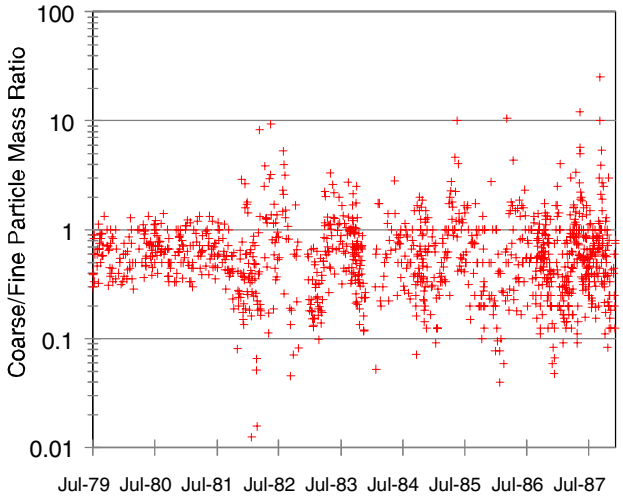

Figure 2. Time-series data from Reconstruct.file for Portage 1979–1987. The scattering of data points shows that data from epoch 1 (1979–1981) were systematically excluded whenever the coarse/fine mass ratio was less than 0.3 or greater than 1.3.

Exclusion of Samples Collected from Multiple Filters

Another selection criterion employed was the exclusion of concentrations measured with more than one set of filters. Samples were sometimes collected over multiple filters because the Beckman dichotomous samplers automatically switched to a new filter pair whenever the fine particle

Table 8. Comparability of Dichotomous Sampler Fine Particle Mass Concentrations from Original Records Inventoried in Table 7, Reconstruct.file (NEJM), and JAWMA.file for the Six Cities Study

	Period Audited		Number of Values			Number Unmatched[a]				Mean Value (μg/m^3)		
City	Start	End	Original Records	NEJM	JAWMA	Original Records vs NEJM	NEJM vs Original Records	Original Records vs JAWMA	JAWMA vs Original Records	Original Records	NEJM	JAWMA
Harriman	01/30/80	05/06/80	69	51	67	18	0	2	0	23.0	22.9	22.5
	02/12/85	12/17/85	205	205	217	0	0	0	12	20.2	20.1	19.7
Portage	02/27/80	07/16/80	68	36	69	32	0	1	2	13.3	14.3	13.3
Steubenville	05/01/80	08/03/80	64	26	60	38	0	4	0	48.0	29.8	46.6
	01/19/84	07/11/84	84	82	84	2	0	0	0	26.1	26.6	26.1
St Louis	03/20/80	09/08/80	97	70	96	27	0	1	0	23.7	25.3	23.5
	03/10/82	04/01/82	12	12	12	0	0	0	0	14.3	13.8	13.8
	05/12/83	08/01/83	30	30	30	0	0	0	0	22.2	22.8	22.8
	02/18/85	01/01/86	157	153	173	4	0	3	19	17.5	17.9	17.6
Topeka	02/21/84	08/01/84	96	96	97	1	1	0	1	12.6	12.7	12.6
	02/03/85	05/22/85	72	72	76	0	0	0	4	10.4	10.4	11.0
Watertown	01/10/85	12/31/85	203	177	210	28	2	19	26	14.2	14.6	14.9
Totals			1,157	1,010	1,191	150	3	30	64			

[a] The number of entries in the first file for which no corresponding entries were found in the second.

Table 9. Effects of Excluding Observations Outside the Acceptable Range of Coarse/Fine Particle Mass Ratio (0.3–1.3) in the Six Cities Study[a]

	Number of Observations		Average Fine Particle Concentration (μg/m^3)		
City	All Data[b]	Restricted[c,d]	All Data[b]	Restricted[c]	Percentage of Change[e]
Harriman	699	546 (78%)	19.6	19.2	–2
Portage	508	310 (61%)	10.5	10.5	0
Steubenville	541	424 (78%)	26.1	26.6	2
St Louis	588	443 (75%)	17.8	18.0	1
Topeka	557	270 (48%)	11.7	13.5	15
Watertown	602	399 (66%)	14.1	14.1	0

[a] All data in the Reconstruct.file for 1982–1985.

[b] Values for all days (observations) regardless of the coarse/fine mass ratio value.

[c] Values for days (observations) on which the coarse/fine mass ratio fell within the range.

[d] Percentage of all observations in parentheses.

[e] This was calculated as [(Restricted – Whole)/Restricted] × 100%.

flow dropped below a specified rate (14.25 L/min, from a nominal 16.7 L/min). As Briggs and colleagues (1982) noted, "This can be expected to happen during very polluted days, when the filters become heavily loaded.... [This condition] (multiple samples in a day) does not indicate questionable data." Rejecting these observations could have incorrectly attenuated high concentrations.

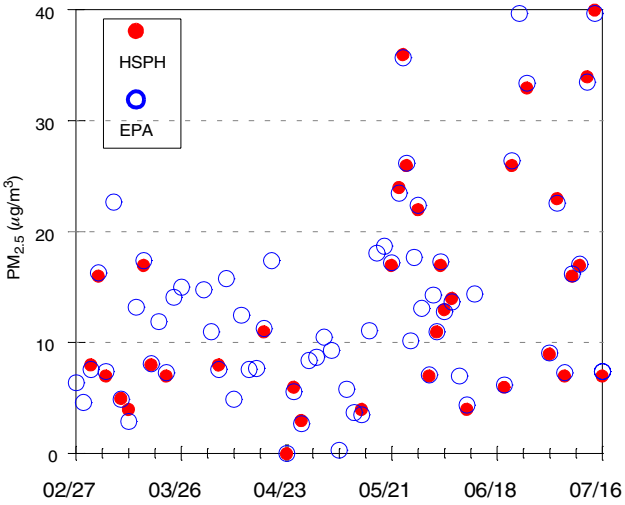

Figure 3. Data for Portage from 2/27/80 to 7/16/80 restricted due to coarse/fine mass ratio. Open circles are data transmitted from the EPA laboratory and filled circles are data from the HSPH Reconstruct.file. Misalignment of coincident EPA and HSPH values reflects numerical rounding of the HSPH values. The EPA data points that have no matching HSPH data points were data excluded from the HSPH files because the coarse/fine mass ratio fell outside the acceptable range of 0.3–1.3. Average fine particle mass was 13.3 μg/m^3 for the 68 EPA measurements and 14.3 μg/m^3 for the 36 HSPH values.

Figure 4 compares the data in the Reconstruct.file with the values reported to HSPH by the EPA laboratory in Steubenville in 1980 that included high concentrations of fine particles. As suggested by Briggs and colleagues (1982), concentrations were generally higher on days when multiple filters were used. The EPA laboratory

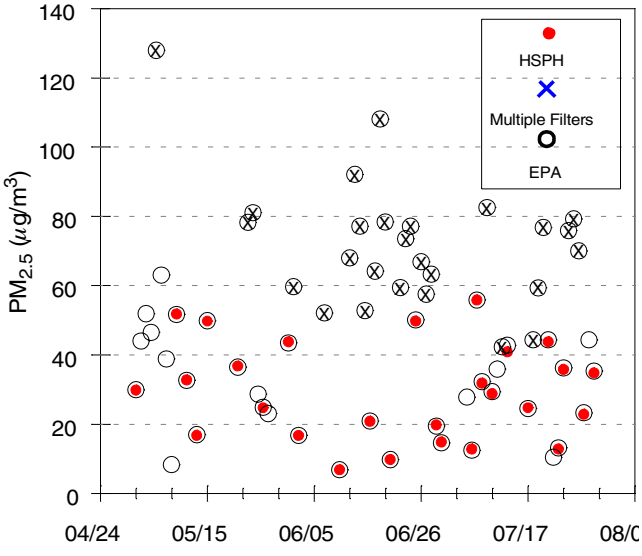

Figure 4. Fine particle levels for a period of time in Steubenville in 1980 that included high concentrations of fine particles. Open circles are values reported to HSPH by the EPA contract laboratory; an ⊗ indicates a value obtained from multiple filters; filled circles are data in the HSPH Reconstruct.file. The recordings of higher concentrations were generally on days when multiple filters were used. Of the 64 EPA laboratory observations, 26 measurements had been acquired with multiple filters and were excluded from the HSPH analysis. The HSPH Reconstruct.file reports 26 concentrations, all from single filters.

Table 10. City Mean Particle Concentrations Published in NEJM and Recalculated from Reconstruct.file for Indicated Years of the Six Cities Study[a]

City	Published (NEJM)[b]	Mean of All Observations	Mean of Annual Averages	Mean of Quarterly Averages
Fine Particles (1979–1985)				
Harriman	20.8	20.9	20.8	20.9
Portage	11.0	11.0	11.0	11.0
Steubenville	29.6	29.6	29.7	29.6
St Louis	19.0	19.0	19.7	19.0
Topeka	12.5	12.5	12.9	12.5
Watertown	14.9	14.9	15.2	14.9
Inhalable Particles (1979–1985)				
Harriman	32.5	32.6	32.5	32.6
Portage	18.2	18.2	18.2	18.1
Steubenville	46.5	46.5	46.4	46.4
St Louis	31.4	31.4	33.0	31.3
Topeka	26.4	26.4	26.3	26.4
Watertown	24.2	24.2	24.6	24.1
Total Particles (1977–1985)				
Harriman	49.4	49.4	49.4	49.9
Portage	34.1	33.4	34.1	32.0
Steubenville	89.9	92.4	89.9	91.2
St Louis	72.5	68.7	72.5	68.3
Topeka	56.6	56.2	56.6	54.3
Watertown	49.2	46.6	49.2	46.3

[a] Values are given as means in $\mu g/m^3$.

[b] See Table 1 in Dockery et al 1993.

reported values for 64 observations for which the average fine particle mass concentration was 48 $\mu g/m^3$. Of those 64, 26 measurements had been acquired with multiple filters; with those 26 values eliminated, the fine particle mass concentration was 32 $\mu g/m^3$ for the 38 observations from single filters. The Reconstruct.file reports values for 26 observations [from single filters] for which the average fine particle mass concentration was 30 $\mu g/m^3$. Multiple-filter observations became less frequent in later years.

Reproducing the Published Statistics Table 10 shows mean concentrations for fine, inhalable, and total particles from the NEJM publication and for three different calculations from data in the Reconstruct.file. The first calculation (Mean of All Observations) averages all observations within the indicated time periods. The second calculation (Mean of Annual Averages) represents an average of yearly concentrations. The third calculation (Mean of Quarterly Averages) represents an average of quarterly mean concentrations. The recalculated Means of Annual Averages exactly match the published means for total particles at all six cities. However, the corresponding means for fine and inhalable particles differ significantly from the NEJM values at St Louis, Topeka, and Watertown. Conversely, the recalculated Means of All Observations exactly match or are within 0.1 $\mu g/m^3$ (Harriman) of the NEJM values for fine and inhalable particles, but are significantly different for total particles at all cities except Harriman.

Table 11. City Mean Particle Concentrations Calculated from All Observations for 1979–1985 in Reconstruct.file and in JAWMA.file for the Audit of the Six Cities Study

	Concentration ($\mu g/m^3$)		Number of Observations	
City	Reconstruct[a]	JAWMA	Reconstruct	JAWMA
Fine Particles				
Harriman	20.9	21.0	1,029	1,552
Portage	11.0	11.5	771	975
Steubenville	29.6	30.8	994	1,145
St Louis	19.0	18.9	868	1,046
Topeka	12.5	12.5	728	938
Watertown	14.9	15.7	850	1,139
Inhalable Particles				
Harriman	32.6	33.0	1,026	1,151
Portage	18.2	18.5	737	925
Steubenville	46.5	48.3	987	1,143
St Louis	31.4	31.7	852	1,043
Topeka	26.4	28.3	720	938
Watertown	24.2	24.5	836	1,139

[a] Referred to as the Mean of All Observations in Table 10.

Other Evidence for the Quality of the NEJM Air Pollution Data

Comparison of Reconstruct.file with the JAWMA.file The three recalculated means in Table 10 are all derived from the Reconstruct.file. All three therefore reflect the same selection criteria for the inclusion or exclusion of observations. To understand the effects of altering these criteria, the Audit Team compared the Means of All Observations shown in Table 10 with the same statistics calculated for data in the JAWMA.file (Schwartz et al 1996), the results of which are shown in Table 11. (This comparison does not include total particle concentrations because the JAWMA.file included data from dichotomous samplers only, which provide fine and coarse particle levels.)

The selection criteria used to extract the data in the JAWMA.file were undocumented, but probably were based on less stringent criteria than those used in the Reconstruct.file. Averaging all observations in the JAWMA.file for the 1979–1985 period does not yield the means published in NEJM, even though averaging all observations including those from later years does yield exactly the time-series means published in JAWMA (data

not shown). As noted in Table 8, about 5% of the examined JAWMA data could not be accounted for in the original records we audited.

Comparison of Reconstruct.file Dichotomous Sampler Data with Data from Size-Selective High-Volume Samplers in 1997.file The high-volume samplers with SSIs measured particles only every sixth day (for 24 hours) during the period 1980 through 1986. These samplers directly measured inhalable particles (fine + coarse) and did not separate fine from coarse. The high-volume samplers' SSIs were different from the dichotomous samplers' SSIs in that (1) they were designed for much higher sample flow rates, and (2) they remained at a 15-μm cutpoint, whereas the dichotomous SSIs changed to a 10-μm cutpoint in early 1984.

The data from the high-volume samplers with SSIs had not been used either in the cross-sectional analysis published in NEJM or in the time-series analysis published in JAWMA due to the low frequency of the measurements. The data had, however, been quality assured along with the other particle measurements (Spengler et al 1986). The SSI high-volume sampler had been operated independently from the dichotomous sampler; not only were the particles sized and the airflows controlled separately, but different filter media and analytic procedures had been used. The data from the SSI high-volume samplers could thus be used to corroborate the data from the dichotomous samplers in that agreement between two independent measurements provides evidence of the quality of both sets of measurements. Due to the different sampling schedules, this comparison did not address the issues of data selection and file integrity.

The paragraphs that follow adopt a temporary convention restricting the use of the term "inhalable particles". Previously, we have used the term to refer to any particles having diameters less than 10 or 15 μm. The concentrations of inhalable particles reported in NEJM, in particular, had been derived by adding together the separate concentrations obtained from the dichotomous samplers for fine (diameters < 2.5 μm) and coarse (diameters > 2.5 μm and < 10 or 15 μm) particles. *In this section only*, the term *IP* is reserved for data from the SSI high-volume samplers; data from the dichotomous samplers are distinguished as *FP*, *CP*, and *FP+CP*.

We expect the relation between concentrations from the high-volume SSI (IP_{HV}) and dichotomous (FP_{DC}+CP_{DC}) samplers to follow the form

$$IP_{HV} = a_0 + a_M\,(FP_{DC}+CP_{DC}),$$

Table 12. City-Specific Coefficients for Regressions[a] Calculated for the Six Cities Study

City	n	r^2	a_M (mean + SE)	a_0 (mean + SE)
Harriman	359	0.78	0.95 ± 0.03	8.1 ± 0.9
Portage	283	0.67	0.95 ± 0.04	5.1 ± 0.9
Steubenville	316	0.88	1.02 ± 0.02	9.2 ± 1.0
St Louis	284	0.78	1.04 ± 0.03	9.5 ± 1.1
Topeka	283	0.71	0.95 ± 0.04	10.2 ± 1.1
Watertown	225	0.46	0.83 ± 0.06	11.0 ± 1.7

[a] Regressions took the form $IP_{HV} = a_0 + a_M\,(FP_{DC}+CP_{DC})$, where IP_{HV} is from the 1997.file SSI high-volume sampler data and $FP_{DC}+CP_{DC}$ is from the JAWMA.file dichotomous sampler data. n is the number of observations (number of days) for which values were available from both samplers. The estimated coefficients a_M and a_0 are explained in the text.

where *HV* refers to the concentrations obtained from high-volume samplers and *DC* refers to the concentrations from the dichotomous samplers; the intercept $a_0 > 0$ is a measurement artifact associated with the high-volume sampler filters used in the measurement of inhalable particles; and the coefficient a_M should equal unity if the dichotomous and SSI high-volume sampler measurements are equivalent. The constant a_0 allows for the extra mass in the IP_{HV} samples contributed by artifactual sulfate (discussed in the next section). The ordinary-least-squares coefficients determined by city-specific regression of the 1997.file SSI high-volume sampler IP_{HV} data on the JAWMA.file dichotomous $FP_{DC}+CP_{DC}$ data are summarized in Table 12.

Watertown stands out in Table 12 as the city with the weakest correlation (r^2) between IP_{HV} and $FP_{DC}+CP_{DC}$, and as the only city for which the proportionality coefficient a_M differs significantly from 1. The distinctive character of the Watertown measurements is also evident in data plots such as Figure 5. Each point in the plot represents a pair of measurements at the same time and same place, one by SSI high-volume sampler (IP_{HV}) and one by dichotomous sampler ($FP_{DC}+CP_{DC}$). Measurements that agree with each other fall near the diagonal line from lower left to upper right. A relatively large fraction of the Watertown observations lie farther from this line than do the measurements for the other five cities.

We investigated whether the observed scatter in the relation between high-volume and dichotomous sampler measurements at Watertown were due to the high-volume sampler measurements or the dichotomous sampler measurements. We had a limited series of measurements at Watertown made by a high-volume sampler with a 10-μm SSI located at the same sampling site as the high-volume sampler with the 15-μm SSI. We found the 10-μm SSI measurements to be more highly correlated

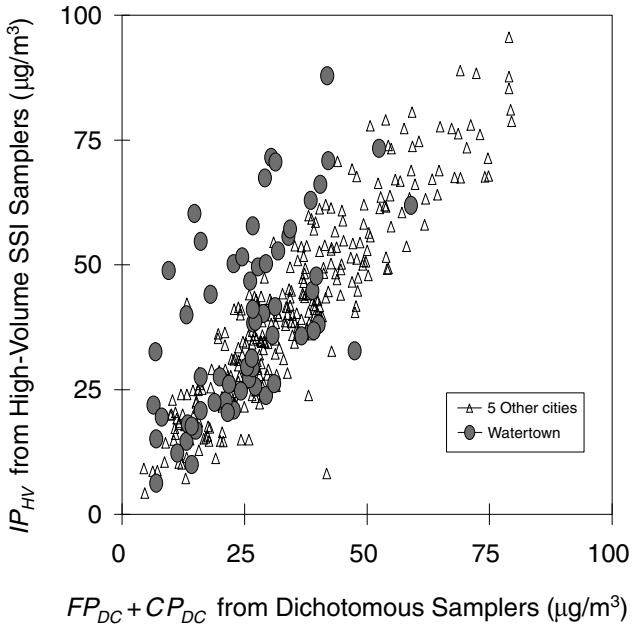

Figure 5. Comparison of particle concentrations gathered in 1980–1981 by different types of samplers. Each point in the plot represents a pair of measurements at the same time and place, one by an SSI high-volume sampler (IP_{HV}) and one by a dichotomous sampler (FP_{DC}+CP_{DC}). Measurements that agree with each other fall near the diagonal line from lower left to upper right. The correspondence between IP_{HV} and FP_{DC}+CP_{DC} measurements in Watertown is noticeably poorer than it is in the other five cities, as evidenced by the relatively large fraction of Watertown observations that lie farther from this line than measurements for the other cities.

with the 15-µm SSI high-volume sampler values than with either the high-volume or the dichotomous sampler measurements. Therefore, we concluded that whatever errors might have occurred in either of the high-volume sampler measurements, they were small compared with the measurement errors in the dichotomous sampler measurements. Furthermore, field logs indicated that the Watertown dichotomous sampler experienced more operational problems and was serviced by more operators than samplers in the other five cities, which supports the contention that the dichotomous sampler measurements were the source of the anomalous values at Watertown.

The SSI high-volume sampler data can also be examined for evidence of the effects from the changes in the dichotomous sampler configurations and the filter analysis methods that differentiated the measurement epochs (described in Table 6). (The SSI high-volume sampler filters were always weighed, whereas the dichotomous filters were sometimes weighed and sometimes analyzed by β attenuation; therefore, the high-volume sampler filter measurements offer a stable reference against which to compare the dichotomous filter measurements.) Because of the anomalous scatter noted above, Watertown has been excluded from this analysis. Table 12 suggests the high-volume sampler artifact (the a_0 column) varies with city but the incremental sensitivity to dichotomous mass (the a_M column) does not. We expect the effects of the changes in sampler configurations and methods to take the approximate form:

$$IP = a_{city} + a_{FP}FP_{DC} + a_{CP}CP_{DC} + d_{grav}(FP_{DC_2} + CP_{DC_2} + b_2) + d_{loss}CP_{DC_{3,4}} + d_{inlet}CP_{DC_4},$$

where FP_{DC_k}, CP_{DC_k}, and b_k take the values FP_{DC}, CP_{DC}, and b during the k^{th} measurement epoch and 0 otherwise. The quantity a_{city} is a city-specific value for the artifact term, a_0. The baseline coefficients a_{FP} and a_{CP} describe the relation of high-volume sampler mass measurements to dichotomous sampler mass measurements during epoch 1, when the dichotomous samplers were operated with unoiled coarse filters and 15-µm inlets and the filters were analyzed by β-absorption gauge. The correction d_{loss} is $-F/(1-F)$, where F is the fractional loss from unoiled coarse filters. Similarly, d_{inlet} is $(R^{-1} - 1)$, where $R = CP_{10}/CP_{15}$ is the ratio of the two definitions of CP mass. Finally, d_{grav} and b describe potential calibration and blank differences between the gravimetric and β-absorption gauge analyses.

Regressions were calculated with various subsets of the above model. Out of 1,525 simultaneous SSI high-volume and JAWMA dichotomous sampler datasets collected in the five cities under examination, only 39 were taken during the few months of epoch 3. The near coincidence of coarse filter oiling at the end of epoch 2, which increased measured CP mass, with the switch from a 15-µm to a 10-µm cutpoint at the start of epoch 4, which decreased measured CP mass, therefore confounded the two changes' opposing effects. The oiling term d_{loss} and the 39 epoch-3 observations were accordingly dropped from the regression, leaving the coefficient d_{inlet} to represent the net effect of the transition from epoch 2 to epoch 4. A small, barely significant calibration effect was associated with the gravimetric analysis, but it was confined to fine particles only. The only robust coefficient was d_{inlet}; Table 13 summarizes results from the regression model setting d_{loss}, d_{grav}, and b equal to zero. The high-volume sampler offset a_{city} was then 6.8, 3.6, 10.4, 10.6, and 7.7 µg/m^3 at Harriman, Portage, Steubenville, St Louis, and Topeka, respectively.

Recall that d_{inlet} in Table 13 represents the net of two effects: postsampling losses were cut by oiling the coarse filters at the same time that the largest particles were dropped from sampling by the more restrictive inlet. Thus the baseline CP coefficient a_{CP} = 1.08 was greater than 1

Table 13. Results and Parameter Estimates from Ordinary-Least-Squares Regressions[a] of *IP* on *FP* and *CP*[b] Calculated for the Six Cities Study

		Epoch 1 (n = 325)	Epoch 2 (n = 450)	Epoch 3 (n = 711)
Variable	t	μg(high-volume)/μg(dichotomous)		
***FP* Total**	4.1	0.91	0.91	0.91
CP	3.0	1.08	1.08	1.08
D_{inlet}	5.8			0.19
***CP* Total**		1.08	1.08	1.27
e_{IP} (μg/m^3)		7.7	7.5	9.7

[a] $R^2 = 0.84$.

[b] *IP* values are from the 1997.file SSI high-volume sampler data, and *FP* and *CP* values are from the JAWMA.file dichotomous sampler data. Student t values are based on standard errors from classical theory; coefficients for *FP* and *CP* are tested against unity, and the adjustment D_{inlet} is tested against zero. *IP* prediction error (e_{IP}) is the root-mean-square difference between observed and predicted *IP* concentrations during indicated measurement epochs.

even in the early years, when the high-volume and dichotomous filters sampled the same particle size range, because it had to account for coarse particles retained by the high-volume sampler filters but lost from the unoiled dichotomous sampler filters. In the absence of such losses, random error in the dichotomous sampler measurements would be expected to attenuate a_{CP} to a value less than 1 in the same way it attenuates a_{FP}.

Comparison of Sulfate Concentrations Measured with High-Volume Samplers and Those Measured with Dichotomous Samplers The sulfate particle data used in the original investigation came from analyses of the high-volume sampler filters. Sulfate particle concentrations were also determined by x-ray fluorescence of the fine and coarse dichotomous sampler filters during the years 1979–1981 and 1984–1988. Table 14 summarizes city-specific regression coefficients between the high-volume and dichotomous determinations for coincident samples (1979–1984). The dichotomous values represent inhalable (fine + coarse) particles; as indicated in the right-hand column, the bulk of inhalable sulfate is in the fine particle fraction. Even perfectly accurate high-volume and dichotomous sampler sulfate values need not be identical because high-volume samples, but not dichotomous samples, could include sulfate carried by "non-inhalable" particles larger than 10 or 15 μm in diameter. The effect of this discrepancy in sampled size ranges is expected to be tiny, however, because the dichotomous samplers found little sulfate in particles larger than 2.5 μm. Figure 6 depicts the correlation between dichotomous and high-volume sampler levels of sulfate for each of the six cities.

Standard high-volume sampler filters are known to react with ambient sulfur dioxide, yielding some artifactual sulfate (Coutant 1977). The Teflon filters used by the dichotomous samplers are inert, thus avoiding this artifact. The expected relation of sulfate measurement from dichotomous samplers to high-volume samplers is thus of the approximate form

$$\text{dichotomous } SO_4^{2-} = b(\text{high-volume } SO_4^{2-} - a),$$

Table 14. Sulfate Concentrations from High-Volume and Dichotomous[a] Particle Samplers for the Six Cities Study

			Regression Coefficients[b]		1979–1984 Interpolated Mean Sulfate		
City	n	r^2	Dichotomous ÷ High-Volume	Artifact	Observed High-Volume Sampler Data	Estimated Dichotomous[c] Sampler Data	Percentage from Fine Particle Channel of Dichotomous Samplers
Harriman	334	0.58	1.21	1.6	8.1	7.9	93
Portage	228	0.81	1.23	1.5	5.3	4.7	92
Steubenville	312	0.79	1.24	1.9	12.8	13.5	88
St Louis	217	0.76	1.10	1.0	8.0	7.6	92
Topeka	143	0.88	1.13	0.9	4.8	4.4	94
Watertown	246	0.71	1.23	1.7	6.5	5.9	90

[a] Values for dichotomous samplers represent inhalable (FP + CP) particles.

[b] Model: dichotomous SO_4^{2-} = (dichotomous/high-volume)(high-volume SO_4^{2-} – artifact). Equal error variances are assumed for dichotomous and high-volume sampler measurements.

[c] Calculated from observed high-volume sampler mean sulfate and the relation of dichotomous sampler to high-volume sampler data described in footnote b.

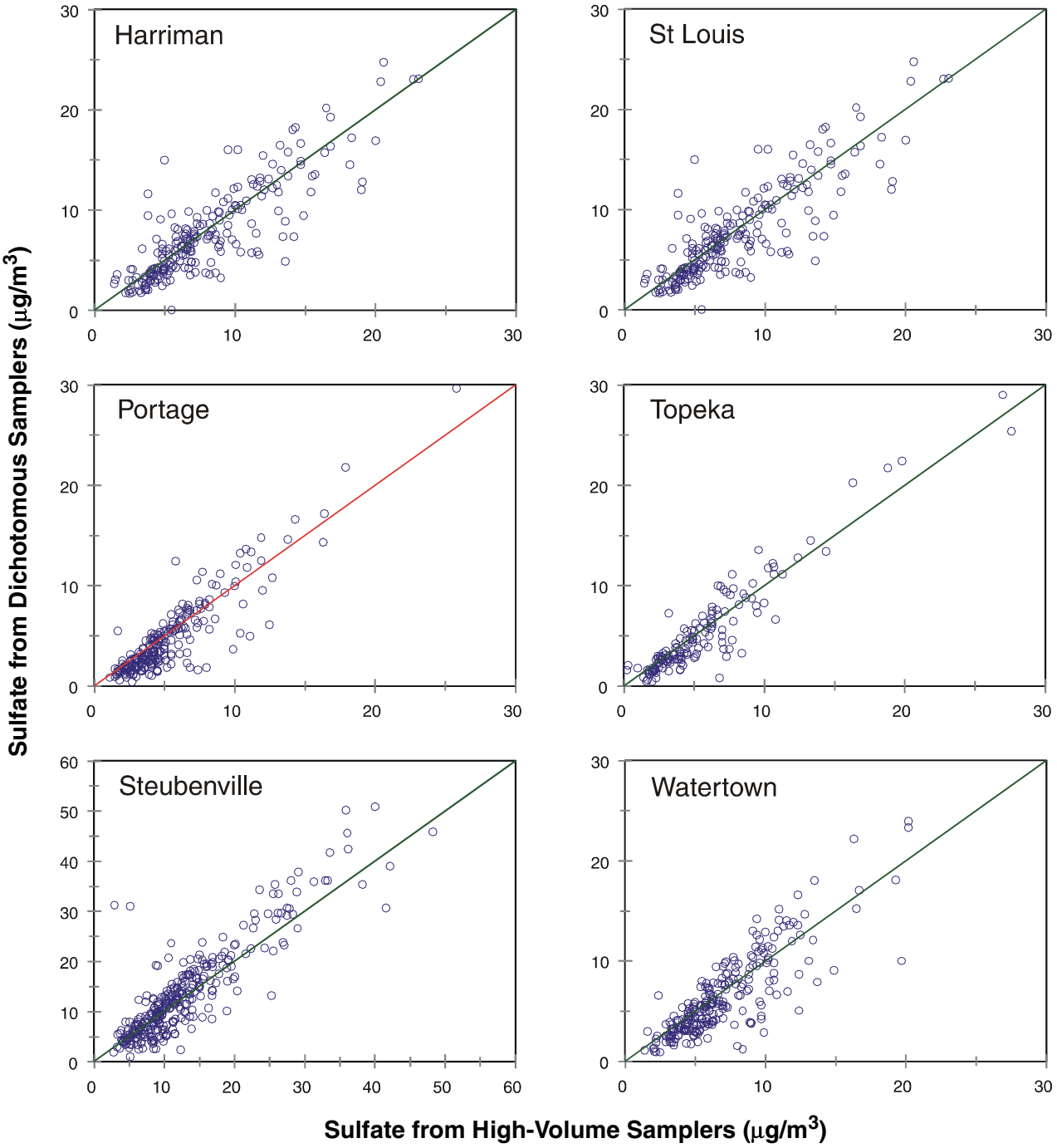

Figure 6. Sulfate determinations by high-volume and dichotomous (fine + coarse) samplers. Note the compressed scale on the Steubenville panel.

where $a \geq 0$ is the characteristic magnitude of the artifact, and $b \leq 1$ is the ratio of incremental dichotomous sulfate to incremental high-volume sulfate. Table 14 gives city-specific coefficients for this relation from least-squares fits with equal weighting of dichotomous and high-volume errors.

The empirical values of high-volume artifact and dichotomous/high-volume slope are reasonably consistent across cities. The apparent artifacts are of plausible magnitude; but the high dichotomous/high-volume slopes are difficult to explain as other than indicators of a systematic error in one of the two determinations. The dichotomous

excess in these coefficients is statistically significant at each city and is evident in data cross-plots.

The empirical relation of *dichotomous* SO_4^{2-} = *b*(*high-volume* SO_4^{2-} – *a*), derived from limited coincident measurements, can be combined with the full data series from the high-volume samplers to estimate the 1979–1984 sulfate averages that would have been obtained from the dichotomous samplers if the 1982–1983 samples had been chemically analyzed. (Ordinary least-squares regression yields essentially the same estimates for 1979–1984 dichotomous sulfate averages, differing by no more than 0.1 μg/m^3.) Table 14 shows the average difference between observed high-volume and estimated dichotomous sulfate to be no more than about 10% because the apparent discrepancy between the high-volume and dichotomous sampler calibrations compensates for the high-volume sampler artifact.

Comparison of Reconstruct.file Total Suspended Particulate Data with the Same Data from 1997.file

The 1997.file includes TSP data for the years 1979–1986, and the NEJM results were based on TSP data for years 1977–1985; therefore, we have restricted our comparisons to the individual years contained in both datasets (1979–1985).

Table 15 summarizes the annual mean TSP concentrations at each city from the 1997.file and from the Reconstruct.file. The column of New TSP concentrations refers to values in the Reconstruct.file but not in the 1997.file. The column of total TSP concentrations

Table 15. Annual Mean TSP and Sulfate Concentrations Calculated from the 1997.file and the Reconstruct.file for the Six Cities Study[a]

	1997.file		Reconstruct.file					
City and Year	TSP	*n*	New TSP[b]	*n*	Total TSP	*n*	SO_4^{2-}	*n*
Harriman								
1979	55.7	170	48.9	60	48.9	60	7.9	170
1980	65.3	177	49.0	61	49.0	61	9.9	177
1981	52.2	166	52.3	39	52.1	193	6.7	166
1982	49.1	170	43.0	1	49.0	171	7.3	170
1983	45.7	131	43.8	5	45.6	136	8.9	131
1984	46.7	122		0	46.7	122	7.8	122
1985	51.5	111		0	51.5	111		
Portage								
1979	36.5	137	38.1	39	36.3	156	5.4	115
1980	33.9	156	38.5	16	34.5	164	5.9	134
1981	31.0	144	25.8	33	30.1	171	4.7	131
1982	33.4	152	22.8	8	32.8	160	5.1	152
1983	33.4	100	25.0	1	33.3	101	5.3	100
1984	31.9	113		0	31.9	113	5.2	113
1985	31.1	111		0	31.1	111		
Steubenville								
1979	77.1	125	104.0	365	104.0	365	14.3	105
1980	74.1	136	90.8	366	90.8	366	13.1	131
1981	74.7	143	86.4	365	86.4	365	13.0	119
1982	66.1	165	82.3	364	82.3	364	12.8	161
1983	63.5	130	72.2	365	72.2	365	12.4	130
1984	70.5	120	72.0	182	72.8	241	11.3	120
1985	68.4	113		0	68.4	113		

(Table continues next page)

[a] Annual mean pollutant concentrations are given in μg/m^3.

[b] Values in the Reconstruct.file but not in the 1997.file.

represents all values in the Reconstruct.file, including the new TSP values. In the early years for all six cities, the TSP data in the 1997.file differ from those in Reconstruct.file; the "early years" vary from city to city, but are easily noted by the presence of data in the "New TSP" column. Note that, for some years at some cities, the data in the Reconstruct.file and 1997.file do not overlap at all. At Harriman, for example, 1997.file had TSP values from every-other-day sampling in 1979–1980, none of which reappear in Reconstruct.file; all of the every-sixth-day TSP values in Reconstruct.file are "new". At Steubenville, similarly, all the the daily TSP values in Reconstruct.file are "new" in the years before 1984.

The TSP values in 1997.file and the sulfate values in Reconstruct.file generally follow a common schedule in each city and appear to have come from the same high-volume sampler, whereas the TSP values in Reconstruct.file appear to include observations from a different instrument. This pattern is evident in Table 16, which compares daily data in the two files for one month at Steubenville. Note that the every-third-day TSP values in 1997.file match those in Reconstruct.file only after July 1; before that, the values for the every-third-day sequence do not match. The Audit Team inferred from this pattern that (1) Reconstruct.file took TSP values after July 1, 1984, from a high-volume that had sampled every third day since 1979 or earlier, whose filters were both weighed for TSP and chemically analyzed for sulfate (SO_4^{2-} in Table 16); (2) Reconstruct.file took TSP values before July 1, 1984, from a different high-volume that had sampled daily until it was taken out of service on July 1, 1984,

Table 15 (*continued*). Annual Mean TSP and Sulfate Concentrations Calculated from the 1997.file and the Reconstruct.file for the Six Cities Study[a]

	1997.file		Reconstruct.file					
City and Year	TSP	*n*	New TSP[b]	*n*	Total TSP	*n*	SO_4^{2-}	*n*
St Louis								
1979	68.0	157	95.8	80	95.4	81	8.6	116
1980	78.9	162	55.8	9	79.5	156	10.4	92
1981	58.0	151	46.2	13	57.1	164	6.7	136
1982	50.5	174		0	50.5	174	7.6	174
1983	50.7	126		0	50.7	126	8.0	126
1984	48.4	117		0	48.4	117	7.5	116
1985	53.9	117		0	53.9	117		
Topeka								
1979	44.7	47	56.3	95	52.5	142	5.2	40
1980	63.8	78	87.3	46	72.6	124	4.4	69
1981	54.4	141	43.3	20	53.0	161	4.5	119
1982	52.8	152		0	52.8	152	4.5	151
1983	54.5	126		0	54.5	126	5.5	126
1984	50.4	114		0	50.4	114	4.9	114
1985	49.5	111		0	49.5	111		
Watertown								
1979	45.9	106	50.9	24	46.8	130	7.4	138
1980	55.7	172	62.2	1	55.7	173	7.7	138
1981	40.8	151	30.1	11	40.1	162	5.1	128
1982	41.3	157		0	41.3	157	6.0	157
1983	39.7	105		0	39.7	105	6.1	105
1984	41.9	108		0	41.9	108	6.5	107
1985	39.3	101		0	39.3	101		

[a] Annual mean pollutant concentrations are given in $\mu g/m^3$.

[b] Values in the Reconstruct.file but not in the 1997.file.

Table 16. Comparison of One Month's Daily Air Pollutant Values for Steubenville from the 1997.file and the Reconstruct.file[a] (Six Cities Study)

	1997.file	Reconstruct.file	
SASDATE	TSP	TSP	SO_4^{2-}
15-Jun-84		53	
16-Jun-84		62	
17-Jun-84	70.6	101	18.4
18-Jun-84		68	
19-Jun-84		62	
20-Jun-84	93.3	66	9.2
21-Jun-84		83	
22-Jun-84		91	
23-Jun-84	63.3	87	16.9
24-Jun-84		63	
25-Jun-84		49	
26-Jun-84	90.4	66	17.8
27-Jun-84		108	
28-Jun-84		90	
29-Jun-84	107.5	92	24.3
30-Jun-84		117	
01-Jul-84			
02-Jul-84	87	87	9.1
03-Jul-84			
04-Jul-84			
05-Jul-84	30	30	7.3
06-Jul-84			
07-Jul-84			
08-Jul-84	69	69	19
09-Jul-84			
10-Jul-84			
11-Jul-84	41.7	41.7	11.7
12-Jul-84			
13-Jul-84			
14-Jul-84	112.4	112.4	3.9
15-Jul-84			

[a] Pollutant values are given in $\mu g/m^3$.

whose filters were weighed for TSP but not chemically analyzed; and (3) Reconstruct.file consistently took sulfate values from the first sampler, thereby taking TSP and sulfate values from separate instruments before July 1, 1984, and from a common instrument after July 1, 1984.

Year-to-year variations in TSP and sulfate provide indirect support for the inference that the early TSP and sulfate data in Reconstruct.file were taken from different high-volume samplers. The notable sulfate maximal levels recorded at Harriman and St Louis in 1980 correspond to the TSP maximal levels in 1997.file that do not appear in Reconstruct.file.

VALIDATION OF THE ORIGINAL SIX CITIES STUDY ANALYSES

We reanalyzed the dataset provided by the Original Investigators by the same methods used in the original analyses by Dockery and colleagues (1993). Specifically, we assessed the effect of air pollution on mortality using the Cox proportional-hazards regression model (Cox 1972). We conducted regression analyses after controlling for the same risk factors considered by the Original Investigators (smoking status, BMI, educational level, and occupational exposure to dusts, gases, and fumes). We stratified all Cox regression models by 5-year age groups and sex, and calculated a baseline hazard for each age-sex group. We used life-table methods to estimate the survival probabilities for each year of follow-up within each city (Cox and Oakes 1983; Lee 1992). The detailed and complete results of the reanalysis of the Six Cities Study data are contained in two appendices that are available from the Health Effects Institute upon request: Appendix E. Computer Programs Used in the Replication of the Original Analyses of the Harvard Six Cities Study; and Appendix F. Replication of the Original Analyses of the Harvard Six Cities Study.

In order to evaluate the reproducibility of the original findings, we summarized the results of the reanalysis in the same format used in the NEJM publication by Dockery and colleagues (1993). Specifically, we compared Tables 1–5 and Figures 1–3 from the publication with the corresponding results of the reanalysis and we provide a description of the findings in the sections that follow.

Validation of the Cohort Selection Process

The Mort6C.file provided by the Original Investigators consisted of a cohort of 8,111 individuals. To replicate the analytic cohort obtained from the Original Investigators, all subjects who completed the initial questionnaire were included. We then selected all individuals who were white, who had two measures of pulmonary function, and whose height was recorded. This cohort consisted of 8,111 individuals and it was identical to the original cohort analyzed by Dockery and colleagues (1993).

Results of the Reanalysis

During the course of the data audit, the Audit Team found that the follow-up of some individuals had been ter-

minated early. Using additional follow-up data provided by the Original Investigators, the Reanalysis Team constructed a second analytic dataset to adust for the problem of early censorship of person-years. When we compared the two cohorts, we discovered that 1% of the members of the original Six Cities cohort had been censored before being lost to follow-up.

The Reanalysis Team conducted two sets of validation analyses for the Six Cities Study. The first analysis was based on the Mort6C.file, which was one version of the Mort6C/HSPH.file used by the Original Investigators. The second analysis was based on the updated analytic cohort that the Reanalysis Team corrected for early censorship.

The results of these two sets of analyses are summarized below. Three versions of each table are shown, labeled as a, b, and c. The first (a) is an exact replica of the table published by Dockery and colleagues (1993); the second (b) presents the results of our validation analysis using the same analytic cohort the Original Investigators had used; and the third (c) presents our results using the updated analytic cohort that we corrected for early censorship. Values presented in bold italic type in the reanalysis tables indicate results different from those reported by the Original Investigators.

Table 17a. Characteristics of the Study Population and Mean Air Pollution Levels in the Six Cities: Original Results[a]

Characteristic	Portage	Topeka	Watertown	Harriman	St Louis	Steubenville
Study Population Variables						
Number of participants	1,631	1,239	1,336	1,258	1,296	1,351
Person-years of follow-up	21,618	16,111	19,882	17,836	17,715	17,914
Number of deaths	232	156	248	222	281	291
Deaths/1,000 person-years	10.73	9.68	12.47	12.45	15.86	16.24
Female sex (%)	52	56	56	54	55	56
Smokers (%)	36	33	40	37	35	35
Former-smokers (%)	24	25	25	21	24	23
Average pack-years of smoking						
Current-smokers	24.0	25.6	25.2	24.5	30.9	28.0
Former-smokers	18.0	19.7	21.8	21.1	22.0	25.0
Less than high school education (%)	25	12	22	35	45	30
Average age (years)	48.4	48.3	48.5	49.4	51.8	51.6
Average body mass index	26.3	25.3	25.5	25.1	26.0	26.4
Job exposure to dust or fumes (%)	53	28	38	50	40	48
Air Quality Variables						
Total particles ($\mu g/m^3$)	34.1	56.6	49.2	49.4	72.5	89.9
Inhalable particles ($\mu g/m^3$)	18.2	26.4	24.2	32.5	31.4	46.5
Fine particles ($\mu g/m^3$)	11.0	12.5	14.9	20.8	19.0	29.6
Sulfate particles ($\mu g/m^3$)	5.3	4.8	6.5	8.1	8.1	12.8
Aerosol acidity ($nmol/m^3$)	10.5	11.6	20.3	36.1	10.3	25.2
Sulfur dioxide (ppb)	4.2	1.6	9.3	4.8	14.1	24.0
Nitrogen dioxide (ppb)	6.1	10.6	18.1	14.1	19.7	21.9
Ozone (ppb)	28.0	27.6	19.7	20.7	20.9	22.3

[a] From Dockery et al 1993; corresponds to Table 1 in the original publication (Copyright © 1993, Massachusetts Medical Society, all rights reserved). Air pollution values were measured in the following years: total particles, sulfur dioxide, nitrogen dioxide, and ozone, 1977 through 1985; inhalable and fine particles, 1979 through 1985; sulfate particles, 1979 though 1984; and aerosol acidity, 1985 through 1988.

Characteristics of the Study Population and Mean Air Pollution Levels in the Six Cities Tables 17a, 17b, and 17c provide a summary of the characteristics of the study population and the air pollution levels in each of the six cities. The study population was characterized according to sex, smoking history, education, age, BMI, and occupational exposure to dust, gases, or fumes. Air pollution was characterized in terms of TSP, inhalable particles, fine particles, sulfate particles, aerosol acidity, sulfur dioxide, nitrogen dioxide, and ozone.

The results of the reanalysis are in close agreement with the original analysis (Table 17b). We found a slight difference in average pack-years of smoking among former-smokers in Watertown; the reanalysis indicated an average of 21.0 pack-years compared with 21.8 pack-years in the original analysis. This appears to be a typographic error in the published results because the original manuscript submitted to NEJM cited the average pack-years of smoking in Watertown as 21.0. The Reanalysis Team also calculated the percentage of participants occupationally exposed to dust, gases, or fumes in Topeka to be 38%, rather than 28% as reported in the original analysis. We also found a few minor differences in estimates of some metrics of particles in Harriman and St Louis. There was a

Table 17b. Characteristics of the Study Population and Mean Air Pollution Levels in the Six Cities: Reanalysis Results Using the Same Analytic Cohort[a]

Characteristic	Portage	Topeka	Watertown	Harriman	St Louis	Steubenville
Study Population Variables						
Number of participants	1,631	1,239	1,336	1,258	1,296	1,351
Person-years of follow-up	21,618	16,111	19,882	***17,835***	17,715	17,914
Number of deaths	232	156	248	222	281	291
Deaths/1,000 person-years	10.73	9.68	12.47	12.45	15.86	16.24
Female sex (%)	52	56	56	54	55	56
Smokers (%)	36	33	40	37	35	35
Former-smokers (%)	24	25	25	21	24	23
Average pack-years of smoking						
Current-smokers	24.0	25.6	25.2	24.5	30.9	28.0
Former-smokers	18.0	19.7	***21.0***	21.1	22.0	25.0
Less than high school education (%)	25	12	22	35	45	30
Average age (years)	48.4	48.3	48.5	49.4	51.8	51.6
Average body mass index	26.3	25.3	25.5	25.1	26.0	26.4
Job exposure to dust or fumes (%)	53	***38***	38	50	40	48
Air Quality Variables						
Total particles ($\mu g/m^3$)	34.1	56.6	49.2	49.4	72.5	89.9
Inhalable particles ($\mu g/m^3$)	18.2	26.4	24.2	***32.6***	31.4	46.5
Fine particles ($\mu g/m^3$)	11.0	12.5	14.9	***20.9***	19.0	29.6
Sulfate particles ($\mu g/m^3$)	5.3	4.8	6.5	8.1	***8.0***	12.8
Aerosol acidity ($nmol/m^3$)	10.5	11.6	20.3	36.1	10.3	25.2
Sulfur dioxide (ppb)	***3.7***	***1.5***	***7.6***	4.8	***9.2***	***23.6***
Nitrogen dioxide (ppb)	6.1	10.6	18.1	14.1	***20.9***	21.9
Ozone (ppb)	28.0	27.6	19.7	20.7	***19.7***	22.3

[a] Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

slightly greater discrepancy in estimates of sulfur dioxide in Portage, Watertown, and St Louis.

Table 17c reports on the characteristics of the study population when the Reanalysis Team eliminated the early censorship of person-years. We found some differences in the person-years of follow-up and the number of deaths reported originally. The person-years of follow-up increased for all six cities; increases ranged from 67 person-years in Watertown to 343 person-years in Portage. The number of deaths increased in Portage (+3), Topeka (+4), Harriman (+2), and Steubenville (+6), and decreased in Watertown (–1).

Adjusted Mortality Rate Ratios Estimated from Cox Proportional-Hazards Models The Cox regression model that we used produced an estimate of the mortality rate, adjusted for age, sex, cigarette consumption, education, and BMI for the six cities. These estimates of risk were relative to Portage. (These are referred to as a mortality rate ratio.) Portage was chosen by the Original Investigators because it had the lowest levels of particles (excluding sulfate particles).

In addition, the Cox model produced estimates of the mortality rate ratio for each of the other variables included in the model (Table 18). For example, the mortality rate for

Table 17c. Characteristics of the Study Population and Mean Air Pollution Levels in the Six Cities: Reanalysis Results After Adjusting for Early Censoring of Person-Years[a]

Characteristic	Portage	Topeka	Watertown	Harriman	St Louis	Steubenville
Study Population Variables						
Number of participants	1,631	1,239	1,336	1,258	1,296	1,351
Person-years of follow-up	***21,961***	***16,342***	***19,949***	***17,911***	***17,789***	***18,052***
Number of deaths	***235***	***160***	***247***	***224***	281	***297***
Deaths/1,000 person-years	***10.70***	***9.79***	***12.38***	***12.51***	***15.80***	***16.45***
Female sex (%)	52	56	56	54	55	56
Smokers (%)	36	33	40	37	35	35
Former-smokers (%)	24	25	25	21	24	23
Average pack-years of smoking						
Current-smokers	24.0	25.6	25.2	24.5	30.9	28.0
Former-smokers	18.0	19.7	***21.0***	21.1	22.0	25.0
Less than high school education (%)	25	12	22	35	45	30
Average age (years)	48.4	48.3	48.5	49.4	51.8	51.6
Average body mass index	26.3	25.3	25.5	25.1	26.0	26.4
Job exposure to dust or fumes (%)	53	***38***	38	50	40	48
Air Quality Variables						
Total particles ($\mu g/m^3$)	34.1	56.6	49.2	49.4	72.5	89.9
Inhalable particles ($\mu g/m^3$)	18.2	26.4	24.2	***32.6***	31.4	46.5
Fine particles ($\mu g/m^3$)	11.0	12.5	14.9	***20.9***	19.0	29.6
Sulfate particles ($\mu g/m^3$)	5.3	4.8	6.5	8.1	***8.0***	12.8
Aerosol acidity ($nmol/m^3$)	10.5	11.6	20.3	36.1	10.3	25.2
Sulfur dioxide (ppb)	***3.7***	***1.5***	***7.6***	4.8	***9.2***	***23.6***
Nitrogen dioxide (ppb)	6.1	10.6	18.1	14.1	***20.9***	21.9
Ozone (ppb)	28.0	27.6	19.7	20.7	***19.7***	22.3

[a] Adjustment for early censoring was based on follow-up through March 15, 1991 for Harriman and through June 30, 1991 for all other cities, as specified by the Original Investigators. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

Table 18a. Adjusted Mortality Rate Ratios Estimated from Cox Proportional-Hazards Models: Original Results for the Six Cities Study[a]

Variable	All Subjects		Men		Women	
Current-smoker	1.59	(1.31–1.92)	1.75	(1.32–2.32)	1.54	(1.16–2.04)
25 Pack-years of smoking	1.26	(1.16–1.38)	1.25	(1.12–1.39)	1.18	(1.00–1.41)
Former-smoker	1.20	(1.01–1.43)	1.17	(0.93–1.48)	1.34	(1.02–1.77)
10 Pack-years of smoking[b]	1.15	(1.08–1.23)	1.16	(1.09–1.25)	1.15	(0.97–1.36)
Less than high school education	1.19	(1.06–1.33)	1.22	(1.06–1.41)	1.13	(0.95–1.35)
Body mass index	1.08	(1.02–1.14)	1.03	(0.95–1.12)	1.11	(1.03–1.20)
City						
Portage[c]	1.00	(—)	1.00	(—)	1.00	(—)
Topeka	1.01	(0.82–1.24)	1.04	(0.79–1.36)	0.97	(0.71–1.34)
Harriman	1.17	(0.97–1.41)	1.21	(0.96–1.54)	1.07	(0.79–1.45)
Watertown	1.07	(0.89–1.28)	0.94	(0.73–1.20)	1.22	(0.93–1.61)
St Louis	1.14	(0.96–1.36)	1.15	(0.91–1.44)	1.13	(0.86–1.50)
Steubenville	1.26	(1.06–1.50)	1.29	(1.03–1.62)	1.23	(0.93–1.61)

[a] From Dockery et al 1993; corresponds to Table 2 in the original publication (Copyright © 1993, Massachusetts Medical Society, all rights reserved). Values are rate ratios (95% CIs). Rates have been adjusted for age, sex, and all other variables listed in the table. The rate ratios for body mass index are for an increase of 4.52 (1 SD).

[b] This actually corresponds to 20 pack-years of smoking. The value 10 in this table was a typographical error in the original paper.

[c] City-specific rate ratios are all expressed in relation to Portage.

Table 18b. Adjusted Mortality Rate Ratios Estimated from Cox Proportional-Hazards Models: Reanalysis Results for the Six Cities Study Using the Same Analytic Cohort[a]

Variable	All Subjects		Men		Women	
Current-smoker	1.59	(1.31–1.92)	1.75	(1.32–2.32)	1.54	(1.16–2.04)
25 Pack-years of smoking	1.26	(1.16–1.38)	***1.26***	***(1.13–1.41)***	1.18	***(0.99***–1.41**)**
Former-smoker	1.20	(1.01–1.43)	1.17	(0.93–1.48)	1.34	(1.02–1.77)
20 Pack-years of smoking	***1.16***	***(1.09***–1.23**)**	***1.17***	***(1.10***–1.25**)**	***1.14***	**(**0.97–***1.35***)
Less than high school education	1.19	(1.06–1.33)	1.22	(1.06–1.41)	1.13	(0.95–1.35)
Body mass index	1.08	(1.02–1.14)	1.03	(0.95–1.12)	***1.12***	(1.03–1.20)
City						
Portage[b]	1.00	(—)	1.00	(—)	1.00	(—)
Topeka	1.01	(0.82–1.24)	1.04	(0.79–1.36)	0.97	(0.71–1.34)
Harriman	1.17	(0.97–1.41)	1.21	(0.96–1.54)	1.07	(0.79–1.45)
Watertown	1.07	(0.89–1.28)	0.94	(0.73–1.20)	1.22	(0.93–1.61)
St Louis	1.14	(0.96–1.36)	1.15	(0.91–1.44)	1.13	(0.86–1.50)
Steubenville	1.26	(1.06–1.50)	1.29	(1.03–1.62)	1.23	(0.93–1.61)

[a] Values are rate ratios (95% CIs). Rates have been adjusted for age, sex, and all other variables listed in the table. The rate ratios for body mass index are for an increase of 4.52 (1 SD). Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] City-specific rate ratios are all expressed in relation to Portage.

Table 18c. Adjusted Mortality Rate Ratios Estimated from Cox Proportional-Hazards Models: Reanalysis Results for the Six Cities Study After Adjusting for Early Censoring of Person-Years[a]

Variable	All Subjects		Men		Women	
Current-smoker	***1.61***	**(*1.33–1.95*)**	***1.77***	**(*1.34–2.34*)**	***1.56***	**(*1.18–2.06*)**
25 Pack-years of smoking	1.26	**(*1.15–1.37*)**	1.25	(1.12–***1.40***)	1.18	**(*0.99*–1.41)**
Former-smoker	***1.21***	**(*1.02*–1.43)**	1.17	**(*0.92–1.47*)**	***1.37***	**(*1.04–1.80*)**
20 Pack-years of smoking	1.15	(1.08–1.23)	***1.17***	(1.09–1.25)	***1.14***	**(*0.96–1.34*)**
Less than high school education	1.19	**(*1.07*–1.33)**	***1.23***	**(**1.06–***1.42*)**	***1.14***	**(*0.96–1.36*)**
Body mass index	1.08	(1.02–1.14)	1.03	(0.95–1.12)	***1.12***	**(*1.04–1.21*)**
City[b]						
Portage	1.00	(—)	1.00	(—)	1.00	(—)
Topeka	***1.03***	**(*0.84–1.26*)**	***1.09***	**(*0.83–1.42*)**	***0.96***	**(*0.70–1.31*)**
Harriman	***1.19***	**(*0.99–1.43*)**	***1.24***	**(*0.98–1.57*)**	1.07	**(**0.79–***1.44*)**
Watertown	1.07	**(**0.89–***1.29*)**	0.94	(0.73–1.20)	1.22	(0.93–1.61)
St Louis	***1.15***	**(**0.96–***1.37*)**	***1.16***	**(*0.93–1.46*)**	***1.12***	**(*0.84–1.47*)**
Steubenville	***1.28***	**(*1.08–1.52*)**	***1.30***	**(*1.04–1.63*)**	***1.26***	**(*0.96–1.66*)**

[a] Values are rate ratios (95% CIs). Rates have been adjusted for age, sex, and all other variables listed in the table. The rate ratios for body mass index are for an increase of 4.52 (1 SD). Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] City-specific rate ratios are all expressed in relation to Portage.

subjects with less than high school education was divided by the mortality rate for those with high school education or more, and this had the value of 1.19 meaning that there was a 19% increase in mortality among the less educated relative to the more highly educated. The adjusted mortality rate ratios are summarized in Tables 18a, 18b, and 18c. In the original version of this table (Table 18a), mortality rate ratios are reported for subjects with 25 and 10 pack-years of smoking. During the course of the reanalysis, we discovered that the rate ratios given for 10 pack-years of smoking actually corresponded to 20 pack-years instead of 10 (Table 18b). We confirmed this with the Original Investigators; it appears this discrepancy was due to a typographic error in the NEJM article.

When the Reanalysis Team adjusted for early censoring of person-years, we found some small changes in the mortality rate ratios (Table 18c); although small, the changes are almost all in an upward direction.

Adjusted Mortality Rate Ratios for the Most-Polluted City Versus the Least-Polluted City Studied Tables 19a, 19b, and 19c present adjusted mortality rate ratios for the most-polluted versus least-polluted city using fine particles as the indicator of air pollution (ie, the mortality rate ratio was calculated for an increase in fine particle concentrations across the range of values represented by the cities; thus, subjects in Steubenville were all assigned a value of 29.6 $\mu g/m^3$, those in Portage 11.0 $\mu g/m^3$, those in Topeka 12.5 $\mu g/m^3$, those in Watertown 14.9 $\mu g/m^3$, those in Harriman 20.9 $\mu g/m^3$, those in St Louis 19.0 $\mu g/m^3$; and

Table 19a. Adjusted Mortality Rate Ratios for the Most-Polluted Versus Least-Polluted Cities Studied by Smoking Status, Sex, and Occupational Exposure: Original Results for the Six Cities Study[a]

Group of Subjects	Number of Subjects	Number of Deaths	Rate Ratio (95% CI)
All	8,096	1,429[b]	1.26 (1.08–1.47)
Nonsmokers	3,266	431	1.19 (0.90–1.57)
Women	2,280	292	1.15 (0.82–1.62)
Men	986	139	1.29 (0.80–2.09)
Former-smokers	1,934	432	1.35 (1.02–1.77)
Women	670	106	1.48 (0.82–2.66)
Men	1,264	326	1.31 (0.96–1.80)
Current-smokers	2,896	566	1.32 (1.04–1.68)
Women	1,478	201	1.23 (0.83–1.83)
Men	1,418	365	1.42 (1.05–1.92)
No occupational exposure[c]	4,455	686	1.17 (0.93–1.47)
Women	3,151	417	1.13 (0.85–1.50)
Men	1,304	269	1.27 (0.85–1.92)
Occupational exposure[c]	3,641	743	1.35 (1.10–1.65)
Women	1,277	182	1.32 (0.86–1.50)
Men	2,364	561	1.35 (1.07–1.69)

[a] From Dockery et al 1993; corresponds to Table 3 in the original publication (Copyright © 1993, Massachusetts Medical Society, all rights reserved). Fine particle concentration was used as the indicator of air pollution. The highest pollution level was in Steubenville and the lowest in Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index. Fifteen subjects were excluded because of missing data on weight.

[b] Although Table 17a indicates a total of 1,430 deaths, the 15 excluded subjects (noted in footnote a) included one death.

[c] To gases, fumes, or dust.

Table 19b. Adjusted Mortality Rate Ratios for the Most-Polluted City Versus Least-Polluted City Studied by Smoking Status, Sex, and Occupational Exposure: Reanalysis Results for the Six Cities Study Using the Same Analytic Cohort[a]

Group of Subjects	Number of Subjects	Number of Deaths	Rate Ratio (95% CI)
All	8,096	1,429	1.26 (1.08–1.47)
Nonsmokers	***3,265***	431	1.19 (0.90–1.57)
Women	2,280	292	1.15 (0.82–1.62)
Men	***985***	139	1.29 (0.80–2.09)
Former-smokers	1,934	432	1.35 (1.02–1.77)
Women	670	106	1.48 (0.82–2.66)
Men	1,264	326	1.31 **(**0.96–***1.79*****)**
Current-smokers	***2,897***	566	1.32 (1.04–1.68)
Women	1,478	201	1.23 (0.83–1.83)
Men	***1,419***	365	1.42 (1.05–1.92)
No occupational exposure[b]	4,455	686	1.17 (0.93–1.47)
Women	3,151	417	1.13 (0.85–1.50)
Men	1,304	269	1.27 (0.85–1.92)
Occupational exposure[b]	3,641	743	1.35 (1.10–1.65)
Women	1,277	182	1.32 **(**0.86–***2.04*****)**
Men	2,364	561	1.35 (1.07–1.69)

[a] Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] To gases, fumes, or dust.

Table 19c. Adjusted Mortality Rate Ratios for the Most-Polluted City Versus Least-Polluted City Studied by Smoking Status, Sex, and Occupational Exposure: Reanalysis Results for the Six Cities Study After Adjusting for Early Censoring of Person-Years[a]

Group of Subjects	Number of Subjects	Number of Deaths	Rate Ratio (95% CI)
All	8,096	***1,443***	***1.28* (*1.10–1.48*)**
Nonsmokers	***3,265***	***433***	***1.22* (*0.92–1.60*)**
Women	2,280	***293***	***1.21* (*0.86–1.70*)**
Men	***985***	***140***	***1.26* (*0.78–2.04*)**
Former-smokers	1,934	***435***	***1.33* (*1.01–1.75*)**
Women	670	***107***	***1.54* (*0.86–2.75*)**
Men	1,264	***328***	***1.28* (*0.94–1.75*)**
Current-smokers	***2,897***	***575***	***1.34* (*1.06–1.70*)**
Women	1,478	***205***	***1.25* (*0.85–1.85*)**
Men	***1,419***	***370***	***1.43* (*1.06–1.93*)**
No occupational exposure[b]	4,455	***694***	***1.21* (*0.96–1.53*)**
Women	3,151	***421***	***1.19* (*0.90–1.58*)**
Men	1,304	***273***	***1.29* (*0.86–1.93*)**
Occupational exposure[b]	3,641	***749***	***1.34* (**1.10–*1.64*)**
Women	1,277	***184***	***1.33* (**0.86–*2.04*)**
Men	2,364	***565***	***1.33* (*1.06–1.67*)**

[a] Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] To gases, fumes, or dust.

the regression model included a term for fine air particles instead of the variable representing cities). The rate ratios are reported by smoking status, sex, and occupational exposure, and are adjusted for age, sex, smoking, education, and BMI. The reanalysis (Table 19b) indicated that the total number of male current-smokers should be 1,419 rather than 1,418 as reported in the NEJM article. The Original Investigators explained that information on weight was missing for one male smoker, so that subject had not been used in this analysis. The Reanalysis Team found an apparent discrepancy in the 95% upper confidence limit on the mortality rate ratio for occupational exposure to gases, fumes, or dust among women; the reanalysis produced an upper limit of 2.04 compared with the original value of 1.50 (Table 19b).

Again, when the Reanalysis Team eliminated the early censorship of person-years (Table 19c), some slight changes in the mortality rate ratios resulted. We did not consider these changes to be of epidemiologic importance.

Estimated Mortality Rate Ratios for the Most-Polluted City Versus the Least-Polluted City in Selected Analytic Models In Tables 20a, 20b, and 20c different analytic models are applied to calculate the mortality rate ratios for the most-polluted city (Steubenville) versus the least-polluted city (Portage) using fine particles as the indicator of air pollution. All rate ratios are adjusted for age and sex. The reanalysis produced results identical to those reported by the Original Investigators (Table 20b).

When the Reanalysis Team corrected for early censorship of person-years, some slight changes were found in all

Table 20a. Estimated Mortality Rate Ratios for the Most-Polluted City Versus the Least-Polluted City Using Selected Models: Original Results for the Six Cities Study[a]

Model Number	Variables Included[b]	Rate Ratio (95% CI)
1	Fine particles	1.31 (1.13–1.52)
2	Model 1 + all smoking variables	1.29 (1.11–1.49)
3	Model 2 + high school education	1.26 (1.08–1.47)
4	Model 3 + body mass index	1.26 (1.08–1.47)
5	Model 4 + occupational exposures	1.26 (1.08–1.46)
6	Model 5 excluding 1,439 subjects with hypertension	1.25 (1.04–1.50)
7	Model 5 excluding 561 subjects with diabetes	1.29 (1.09–1.52)

[a] From Dockery et al 1993; corresponds to Table 4 in the original publication (Copyright © 1993, Massachusetts Medical Society, all rights reserved). Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. In addition to the variables specified, rates have been adjusted for age and sex.

[b] Subjects with hypertension had been treated for high blood pressure within 10 years before enrollment; subjects with diabetes were those who had ever been told by a doctor that they had diabetes, had glucose in their urine, or had too much glucose in their blood.

Table 20b. Estimated Mortality Rate Ratios for the Most-Polluted City Versus the Least-Polluted City Using Selected Models: Reanalysis Results for the Six Cities Study Using the Same Analytic Cohort[a]

Model Number	Variables Included[b]	Rate Ratio (95% CI)
1	Fine particles	1.31 (1.13–1.52)
2	Model 1 + all smoking variables	1.29 (1.11–1.49)
3	Model 2 + high school education	1.26 (1.08–1.47)
4	Model 3 + body mass index	1.26 (1.08–1.47)
5	Model 4 + occupational exposures	1.26 (1.08–1.46)
6	Model 5 excluding 1,439 subjects with hypertension	1.25 (1.04–1.50)
7	Model 5 excluding 561 subjects with diabetes	1.29 (1.09–1.52)

[a] Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. In addition to the variables specified, rates have been adjusted for age and sex.

[b] Subjects with hypertension had been treated for high blood pressure within 10 years before enrollment; subjects with diabetes were those who had ever been told by a doctor that they had diabetes, had glucose in their urine, or had too much glucose in their blood.

Table 20c. Estimated Mortality Rate Ratios for the Most-Polluted City Versus the Least-Polluted City Using Selected Models: Reanalysis Results for the Six Cities Study After Adjusting for Early Censoring of Person-Years[a]

Model Number	Variables Included[b]	Rate Ratio (95% CI)
1	Fine particles	***1.32*** (***1.14–1.54***)
2	Model 1 + all smoking variables	***1.30*** (***1.12–1.51***)
3	Model 2 + high school education	***1.28*** (***1.10–1.48***)
4	Model 3 + body mass index	***1.28*** (***1.10–1.48***)
5	Model 4 + occupational exposures	***1.27*** (***1.10–1.48***)
6	Model 5 excluding 1,439 subjects with hypertension	***1.28*** (***1.07–1.53***)
7	Model 5 excluding 561 subjects with diabetes	***1.30*** (***1.11–1.53***)

[a] Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. In addition to the variables specified, rates have been adjusted for age and sex. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] Subjects with hypertension had been treated for high blood pressure within 10 years before enrollment; subjects with diabetes were those who had ever been told by a doctor that they had diabetes, had glucose in their urine, or had too much glucose in their blood.

Table 21a. Adjusted Mortality Rate Ratios for Current-Smokers, Former-Smokers, and for Both Smoking Groups in the Most-Polluted City Versus the Least-Polluted City by Cause of Death: Original Results for the Six Cities Study[a]

Cause of Death	Percent of Total Deaths	Current-Smoker[b]	Former-Smoker[c]	Most- vs Least-Polluted City
All	100	2.00 (1.51–2.65)	1.39 (1.10–1.75)	1.26 (1.08–1.47)
Lung cancer	8.4	8.00 (2.97–21.6)	2.54 (0.90–7.18)	1.37 (0.81–2.31)
Cardiopulmonary disease	53.1	2.30 (1.56–3.41)	1.52 (1.10–2.10)	1.37 (1.11–1.68)
All others	38.5	1.46 (0.89–2.39)	1.17 (0.80–1.73)	1.01 (0.79–1.30)

[a] From Dockery et al 1993; corresponds to Table 5 in the original publication (Copyright © 1993, Massachusetts Medical Society, all rights reserved). Values are rate ratios (95% CIs). Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index.

[b] The risk of death for a current-smoker with approximately the average number of pack-years of smoking at enrollment (25 pack-years) compared with that for a nonsmoker.

[c] The risk of death for a former-smoker with approximately the average number of pack-years of smoking at enrollment (20 pack-years) compared with that for a nonsmoker.

of the mortality rate ratios (maximum difference 0.03 [Table 20c]).

Adjusted Mortality Rate Ratios for Current-Smokers, Former-Smokers, and for the Most-Polluted City Versus the Least-Polluted City by Cause of Death Tables 21a, 21b, and 21c show adjusted mortality rate ratios for current-smokers and former-smokers, each compared with nonsmokers, and then both smoker groups residing in the most-polluted city versus those in the least-polluted city (with fine particle concentration being used as the indicator of air pollution). For the former two analyses, these rate ratios are adjusted for age, sex, smoking, education, and BMI. These mortality rate ratios represent risk of death for a current-smoker with 25 pack-years of smoking and a former-smoker with 20 pack-years of smoking (the average pack-years at enrollment for each group) compared with never-smokers. The adjusted mortality rate ratios for current-smokers were estimated by multiplying the risk ratio for current-smokers by the risk ratio for the number of pack-years smoked (25). The rate ratios for former-smokers were calculated in a similar fashion.

The Original Investigator determined 95% confidence intervals (CIs) by using the following formula:

$$95\ \%\ \text{CI for RR (Current-Smoker)} = \exp\{\beta_1 + \beta_2 + \beta_3 \pm 1.96\ [\text{Var}(\beta_1) + \text{Var}(\beta_2) + \text{Var}(\beta_3)]^{1/2}\}$$

where β_1, β_2, and β_3 are the estimates of the logarithm of the relative risk for the indicator variable representing current smoking, number of pack-years of cigarettes smoked, and number of years of smoking, respectively, and with the

Table 21b. Adjusted Mortality Rate Ratios for Current-Smokers, Former-Smokers, and for Both Smoking Groups in the Most-Polluted City Versus the Least-Polluted City by Cause of Death: Reanalysis Results for the Six Cities Study Using the Same Analytic Cohort[a]

Cause of Death	Percent of Total Deaths	Current-Smoker[b]	Former-Smoker[c]	Most- vs Least-Polluted City
All	100	2.00 (***1.74–2.31***)	1.39 (***1.20–1.61***)	1.26 (1.08–1.47)
Lung cancer	8.4	8.00 (***3.85–16.63***)	***2.55*** (***1.12–5.80***)	1.37 (0.81–***2.32***)
Cardiopulmonary disease	53.1	2.30 (***1.88–2.82***)	1.52 (***1.23–1.87***)	1.37 (1.11–1.68)
All others	38.5	1.46 (***1.17–1.82***)	1.17 (***0.93–1.48***)	1.01 (0.79–1.30)

[a] Values are rate ratios (95% CIs). Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] The risk of death for a current-smoker with approximately the average number of pack-years of smoking at enrollment (25 pack-years) compared with that for a nonsmoker.

[c] The risk of death for a former-smoker with approximately the average number of pack-years of smoking at enrollment (20 pack-years) compared with that for a nonsmoker.

Table 21c. Adjusted Mortality Rate Ratios for Current-Smokers, Former-Smokers, and for Both Smoking Groups in the Most-Polluted City Versus the Least-Polluted City by Cause of Death: Reanalysis Results for the Six Cities Study After Adjusting for Early Censoring of Person-Years[a]

Cause of Death	Percent of Total Deaths	Current-Smoker[b]	Former-Smoker[c]	Most- vs Least-Polluted City
All	100	***2.03*** (***1.76–2.33***)	1.39 (***1.20–1.61***)	***1.28*** (***1.10–1.48***)
Lung cancer	***8.2***	***8.07*** (***3.89–16.75***)	***2.52*** (***1.10–5.74***)	***1.43*** (***0.85–2.41***)
Cardiopulmonary disease	***51.7***	2.30 (***1.88–2.82***)	1.52 (***1.23–1.87***)	***1.38*** (***1.12–1.69***)
All others	***37.6***	***1.44*** (***1.16–1.80***)	1.17 (***0.94–1.47***)	1.01 (0.79–1.30)

[a] Values are rate ratios (95% CIs). Fine particle concentration was used as the indicator of air pollution. The city with the highest pollution level was Steubenville and that with the lowest was Portage. Rates have been adjusted for age, sex, smoking, education, and body mass index. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] The risk of death for a current-smoker with approximately the average number of pack-years of smoking at enrollment (25 pack-years) compared with that for a nonsmoker.

[c] The risk of death for a former-smoker with approximately the average number of pack-years of smoking at enrollment (20 pack-years) compared with that for a nonsmoker.

corresponding estimates of variance denoted by Var(β•). Interval estimation using this approach assumes that the parameter estimates are statistically independent, though these parameters are actually correlated.

When recalculating CIs for current- and former-smokers, the Reanalysis Team incorporated statistical dependence between the parameter estimates into the calculation of the CI by applying the formula:

$$95\%\ \text{CI for RR (Current-Smoker)} = \exp(\beta_1 + \beta_2 + \beta_3 \pm 1.96\ \{\text{Var}(\beta_1) + \ldots + \text{Var}(\beta_3) + 2[\text{Cov}(\beta_1, \beta_2) + \ldots + \text{Cov}(\beta_1, \beta_3)]\}^{1/2})$$

where $\text{Cov}(\beta_1,\beta_2)$ is the estimated covariance between the parameter estimates. (We refer to this as a direct method.) Covariances were estimated using the SAS procedure for the Cox proportional-hazards model. The CIs are narrower using this approach than those determined by the method the Original Investigators used (Table 21b).

Once again, when the Reanalysis Team corrected for the early censorship of person-years, we noted slight increases in the risk ratios (Table 21c).

Annual Average Concentrations of Total Particles, Fine Particles, and Sulfate Particles in the Six Cities Figures 7 through 9 show the levels of TSP, fine particles, and sulfate particles in each city. In seeking to validate the original results on the basis of air quality data provided by the Original Investigators, the Reanalysis Team found some discrepancies in what had been published in the NEJM article. The Reanalysis Team received directly from Dr Dockery on July 29, 1999, the dataset we used to recompute the long-term means published by Dockery and colleagues. The dataset was used by the Reanalysis Team to reproduce the long-term averages and annual average concentrations of pollutants cited in the original paper.

The Reanalysis Team noted a number of discrepancies among the published averages, those received from Dr Dockery (personal communication from Douglas Dockery to the Reanalysis Team on March 31, 1999) and the ones we computed. The results of this step are summarized in Tables 17a, 17b, and 17c.

For all gaseous pollutants other than sulfur dioxide, the discrepancies the Reanalysis Team noted were minor and could be attributed to approximations in intermediate steps or to use of different software or procedures within the same software. For St Louis, the mean concentrations for nitrogen dioxide and ozone were apparently reversed in the NEJM article. This was either a typographic or transcription error in the article.

We were not able to reproduce the mean concentrations for sulfur dioxide except for Harriman. The discrepancies in means ranged from 0.1 $\mu g/m^3$ for Topeka to 4.9 $\mu g/m^3$ for St Louis. The published means for both TSP and sulfur dioxide were computed from annual averages (personal communication from Douglas Dockery to the Reanalysis Team on March 31, 1999). The Reanalysis Team followed the same procedures. We calculated annual averages first and then used those

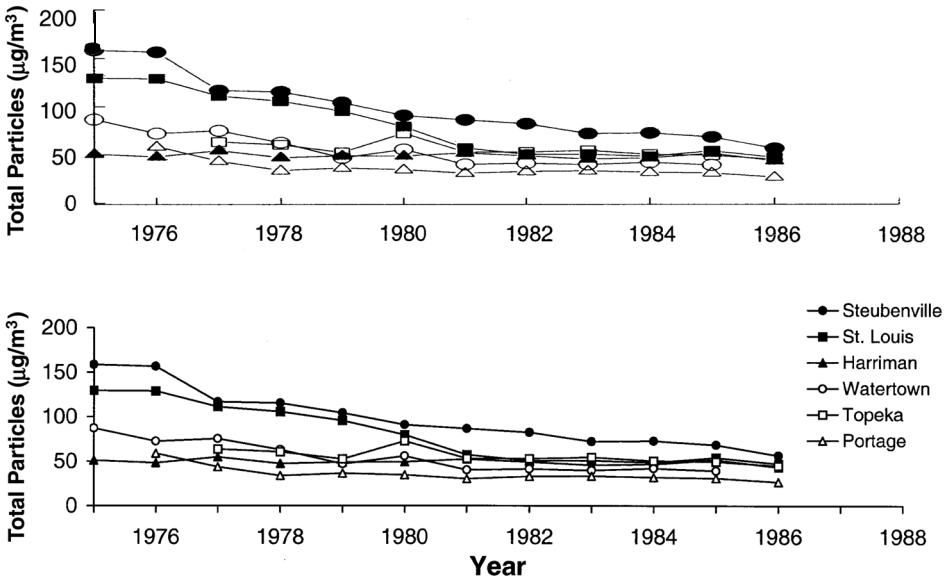

Figure 7. Annual Average Concentrations of Total Particles in the Six Cities. Top panel: Original results from Dockery and associates 1993 (Figure 1 top panel; Copyright © 1993, Massachusetts Medical Society, all rights reserved). Bottom panel: Reanalysis Team's results.

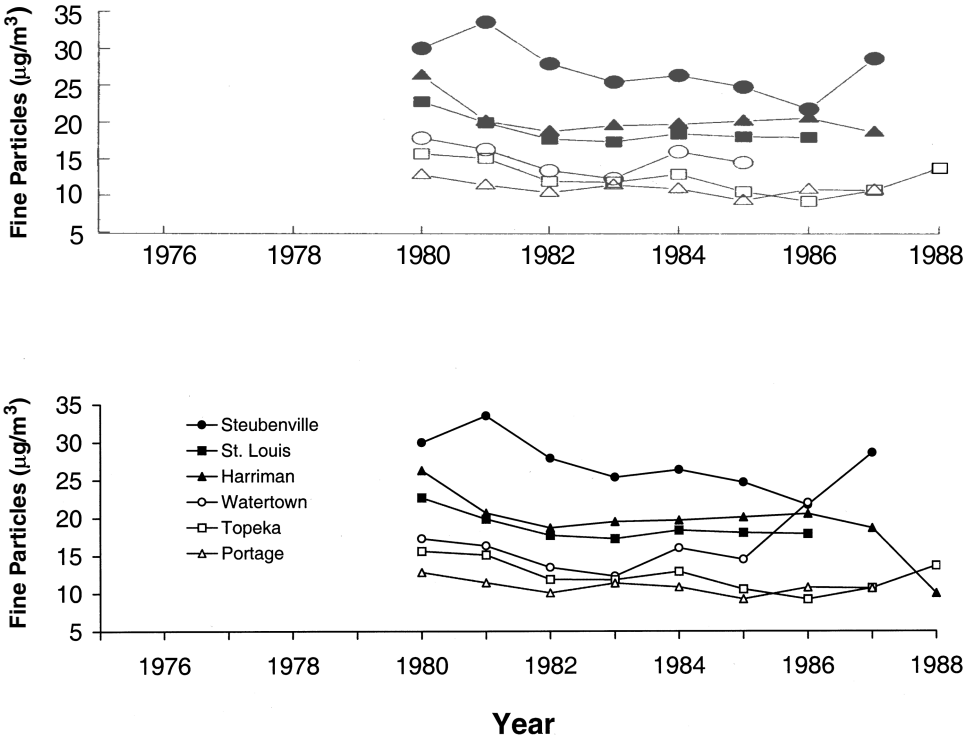

Figure 8. Annual Average Concentrations of Fine Particles in the Six Cities. Top panel: Original results from Dockery and associates 1993 (Figure 1 middle panel; Copyright © 1993, Massachusetts Medical Society, all rights reserved). Bottom panel: Reanalysis Team's results. (Note: The Original Investigators did not use the 1986 data for Watertown or the 1988 data for Kingston-Harriman because only one measurement was taken in these two cities in those years.)

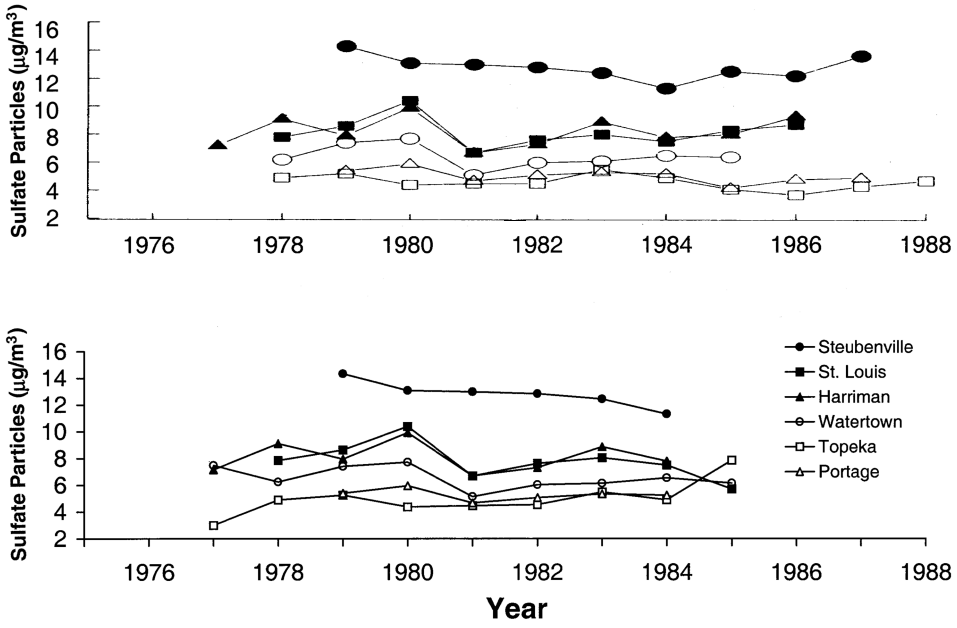

Figure 9. Annual Average Concentrations of Sulfate Particles in the Six Cities. Top panel: Original results from Dockery and associates 1993 (Figure 1 bottom panel; Copyright © 1993, Massachusetts Medical Society, all rights reserved). Bottom panel: Reanalysis Team's results.

to compute long-term average concentrations. However, the discrepancies still persisted.

Some discrepancies are noticeable in the graphic plots as well. Figures 7, 8, and 9 show differences in years of coverage between the published and the computed plots. First, for TSP (Figure 7), data entries prior to 1975 (1973 and 1974 data mostly) were not used in the original publication; from 1975 on, the original and reanalysis plots are consistent. Second, the data points for fine particles (Figure 8) prior to 1980 were omitted from the original graph. These data exist for all cities except Harriman. In addition, fine particle data for later years were not shown in the original plots; these are Harriman in 1988 and Watertown in 1986, as shown in Figure 8. Third, data for sulfate particles before 1978 were not used in the original analyses, except for Harriman, where the data start in 1977 (Figure 9). In the original figure, data are plotted for all cities (except Watertown) beyond 1985. However, the Reanalysis Team found no data entries for the years 1986–1988 in the data file; in fact, we found data for 1985 only for St Louis, Topeka, and Watertown.

Crude Probability of Survival in the Six Cities by Years of Follow-up Figure 10 (which was Figure 2 in NEJM) illustrates the crude probability of survival in each of the six cities according to the number of years of follow-up. The Reanalysis Team found no differences between our results and those reported by the Original Investigators.

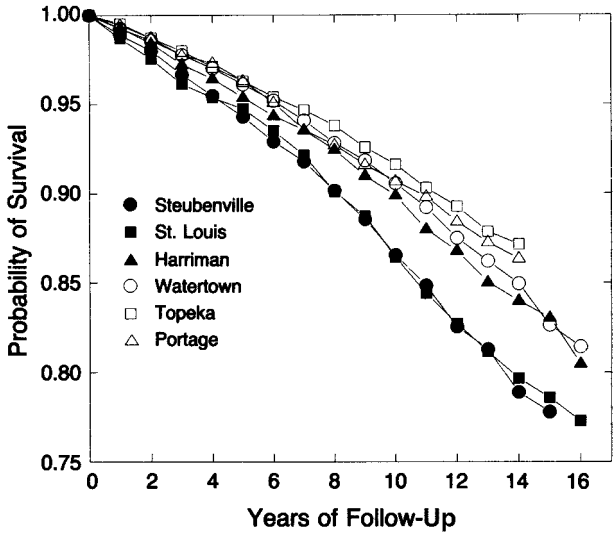

Figure 10. Crude Probability of Survival in the Six Cities, According to Years of Follow-Up. Original results from Dockery and associates 1993 (Figure 2; Copyright © 1993, Massachusetts Medical Society, all rights reserved).

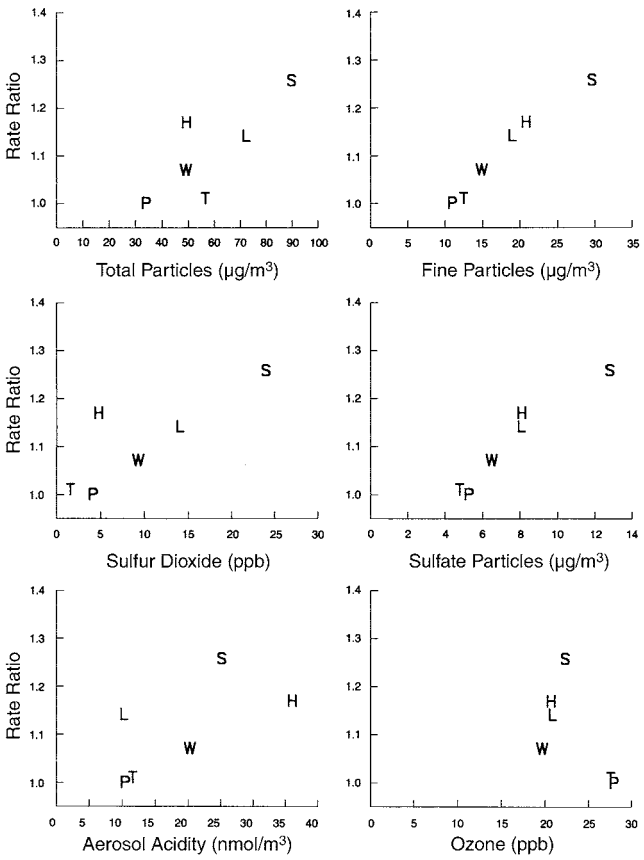

Figure 11. Estimated Adjusted Mortality Rate Ratios and Pollution Levels in the Six Cities. Original results from Dockery and associates 1993 (Figure 3; Copyright © 1993, Massachusetts Medical Society, all rights reserved). P = Portage, T = Topeka, W = Watertown, L = St Louis, H = Harriman, and S = Steubenville.

Estimated Adjusted Mortality Rate Ratios and Pollution Levels in the Six Cities Each panel of Figure 11 (which was Figure 3 in the NEJM) shows the relation between mortality rate ratios in each city on the basis of one measure of air pollution: TSP, fine particles, sulfur dioxide, sulfate particles, aerosole acidity, or ozone. The reanalysis revealed no discrepancies in the original findings.

THE AMERICAN CANCER SOCIETY STUDY

In 1982, the ACS initiated a large prospective cohort study, which involved subjects from all 50 United States, the District of Columbia, and Puerto Rico, known as the Cancer Prevention Study II (CPS-II). Enrollment had been restricted to persons who were at least 30 years of age and who were members of households with at least one individual 45 years of age or older. Each participant had completed a four-page questionnaire (see Appendix D), which included items on age, sex, weight, height, demographic characteristics, family history of cancer, disease history, use of medication and vitamins, occupational exposures, dietary habits, use of alcohol and tobacco, and aspects of exercise and health-related behaviors.

Vital status for all CPS-II participants, from September 1, 1982, through December 31, 1989, had been determined using personal inquiries and automated record linkage to the NDI. Death certificates had been subsequently obtained from state health departments and coded by a nosologist. The nosologist coded the underlying cause of death according to the ICD-9.

From the CPS-II cohort of approximately 1.2 million adults, Pope and colleagues (1995) included all subjects who had no missing data for a specific set of variables obtained from the questionnaire, and for whom a death certificate had been obtained if they were deceased. Two subsets of this population were defined if they had resided, at the time of enrollment, in (1) one of 151 metropolitan areas (MAs) for which sulfate particle measurements had been collected during the years 1980–1982, or (2) one of 50 MAs for which fine particle measurements had been collected during the years 1979–1983. (The 151 MAs with sulfate measurements included all but three of the 50 MAs with fine particle measurements; thus, data were collected for 154 total cities.) The population subset with exposure to sulfate particles totaled 552,138 adult subjects (referred to as the sulfate cohort), of which 295,223 subjects were also in the population subset with exposure to fine particles (referred to as the fine particle cohort). Risk factor data for individuals were obtained from the CPS-II. (Hereafter, the study by Pope and associates [1995], as published in the *American Journal of Respiratory and Critical Care Medicine* [AJRCCM], is referred to as the ACS Study.[†])

A total of 38,963 deaths were recorded in the sulfate cohort and 20,765 deaths in the fine particle cohort. Separate analyses were performed for deaths from: all causes combined, lung cancer, cardiopulmonary disease, and all others.

Two measures of air pollution, fine particles and sulfate, were modeled. The mean concentration of sulfate air pollution by MA during 1980 was estimated using data from the EPA Aerometric Information Retrieval System (AIRS) database. These means were calculated as the averages of annual arithmetic mean 24-hour sulfate values for all monitoring sites in the 151 MAs. Mean sulfate concentrations averaged 11 $\mu g/m^3$ and ranged from 3.6 to 23.5 $\mu g/m^3$. The median concentration of fine particles between 1979 and 1983 was estimated from the

[†] The original article appears in its entirety at the end of this Special Report.

EPA's dichotomous sampler network by Lipfert and colleagues (1988). These estimates of fine particle levels had been used previously in a population-based cross-sectional mortality study of 50 MAs. The average median fine particle concentration was 18.2 $\mu g/m^3$ and overall values ranged from 9.0 to 33.5 $\mu g/m^3$ (Lipfert 1993).

Cox proportional-hazards regression models were used to calculate adjusted mortality risk ratios. The time axis was defined using calendar time from the date of enrollment. Statistical adjustments were made for several covariates that included, among others, smoking, education, BMI, and alcohol consumption. In addition, the potential confounding influence of occupational exposures on the estimates of air pollution, such as diesel engine exhaust, wood dust, and fumes, was evaluated. All models were stratified by 5-year age categories, gender, and race, which allowed each sex-age-race stratum to have its own baseline hazard. To determine the extent to which the results were confounded by differences in climates across the MAs, variables that accounted for relatively hot or cold conditions were added to the models. Cox proportional-hazards models were estimated separately for all causes of death combined and the three cause-of-death subcategories: lung cancer (ICD-9 code 162), cardiopulmonary disease (ICD-9 codes 401–440 and 460–519), and all others.

Both sulfate and fine particle exposures were found to be associated with an excess risk of all-cause mortality. The ratio of the mortality risk for all causes of death for subjects in the most-polluted city relative to those in the least-polluted city was estimated to be 1.15 (95% CI: 1.09–1.22) for the sulfate cohort and 1.17 (95% CI: 1.09–1.26) for the fine particle cohort.

AUDIT OF STUDY POPULATION DATA

Data Provided and Source Documents Accessible for the Data Audit

In the absence of a study protocol, we audited the data against the study methods and results as presented in the publication by Pope and colleagues (1995). Because of space limitations at the ACS offices, most of the Audit Team's activities were conducted offsite.

One of the difficulties the Audit Team faced was that the original staff of the ACS Study who managed the data collection and databases were no longer employed by ACS. Dr Eugenia Calle of the ACS facilitated contacts with Ms Cathy Lally, who had been employed by ACS after the data collection and much of the programming had been completed. (Ms Lally was no longer employed by ACS, but performed work on a periodic consulting basis.) She assisted the Audit Team with issues about coding and programming.

The ACS staff reconstructed SAS datasets corresponding to the analytic files that had been used by the Original Investigators (hereafter, these electronic datasets are referred to collectively as Analytic.files). These datasets contained all of the variables derived from the questionnaires used in the original analysis (see Appendix D), vital status of the participants, and average annual sulfate and particle levels in the cities.

The Audit Team relied on code books, copies of microfilmed records, and printouts of computer programs provided to the Reanalysis Team. This database has continued to be updated for use in other studies. Therefore, the Audit Team worked with the reconstructed version of the database, as it existed at the time of study publication. From discussions the Audit Team had with Ms Lally, it was clear that, while reconstructing the database for transfer to the Reanalysis Team, she had carefully examined the computer programs and quality control process and responded to any issues that she uncovered. This process was important to the audit, but was not formalized.

Sampling the Dataset and Assessing Error Rates in the Original Data: Subsets of Study Population and Deceased Subjects

The original study cohort included 552,138 men and women who filled out questionnaires on health and lifestyle. As in the audit of the Six Cities Study, we randomly selected questionnaires for 250 subjects. The Audit Team coordinator met with the Reanalysis Team and identified variables from the questionnaires to be verified and used in the sensitivity analyses (Table 22).

Records of vital status had been lost when the ACS offices moved from New York to Atlanta. For 44 of the 250 subjects in the audit sample, the Audit Team ascertained vital status from a later American Cancer Society Nutrition Survey; for these subjects, vital status could be positively confirmed because they were alive at a date later than the termination date for follow-up in the cohort used in the ACS Study. For the remaining 206 subjects, the Audit Team ascertained vital status by checking the NDI; in addition, we searched the Social Security Death Index available on the Internet (http://www.ancestry.com).

The total number of deaths used in the ACS Study analyses was 51,137; from this group of deceased subjects, the Audit Team randomly selected a second 250-subject subset of death certificates. The ACS Original Investigators provided the Audit Team with a list containing full names and dates of birth for these 250 deceased subjects.

For each death certificate, the Audit Team's nosologist coded the underlying cause of death according to the ICD-9 and compared it with the code that had been used in

Table 22. List of Variables for Reanalysis Team to Audit and the Criteria for Declaring Errors in the ACS Study

Original Questionnaire Variable	Subvariable	Criteria
Subject identification number		Any difference
Age at enrollment		Any difference
Sex		Any difference
Race		Any difference
Vital status	Death month and year if applicable	Any difference
Survival time from date of enrollment	Survival censored at end of study, 12/31/89	Any difference
Cigarette smoking status	Current and former	Any difference
Pipe and cigar smoking status	Current and former	Any difference
Years smoked	Current and former	Any difference
Cigarettes per day	Current and former	Any difference
Hours per day exposed to passive smoke		Any difference
Height		Any difference
Weight		Any difference
Number of alcoholic drinks per day		Any difference
Education		Any difference
Occupational exposure to asbestos, chemicals/solvents/acids, coal/stone dust, tar/pitch/asphalt, diesel exhaust, formaldehyde		Any difference

original analysis. The Audit Team identified possible cases in which an ICD-9 code for an immediate or contributing cause of death had been used rather than the ICD-9 code for the underlying cause. From the microfilm copies of each of the death certificates, the Audit Team also tracked the notation of the ICD-9 code through its entry into the Analytic.files and noted any transcription errors.

Printouts of the Analytic.files provided to the Reanalysis Team were used to check specific data points for each variable in the subsets of questionnaires and death certificates.

Subset of Study Population: Questionnaires

Questionnaire Variables All original questionnaires and death certificates had been destroyed after filming because of storage space considerations. Questionnaires were found on the microfilm for 249 out of 250 subjects. One microfilm copy of a questionnaire could not be located because the roll and frame numbers were missing. The Audit Team could not determine if this missing record was due to an error in microfilming or in the actual retrieval and data management of the study.

Questionnaire Validation Variable for SID Number We matched each 14-digit SID number from the Analytic.files with the 14-digit identification number on the questionnaire. Errors were found in 3 (1.2%) of the 249 questionnaires: one in the division number, one in the unit number, and one in the group number.

Sex and Race We found no inconsistencies in the recording of sex or race.

Age at Enrollment The Audit Team noted one minor inconsistency in recording the age at enrollment in that the age had been rounded up to the next year.

Height and Weight We detected no errors in the presentation of height and weight.

Smoking Status and Passive Exposure to Smoke Information on active smoking and passive exposure to smoke was contained in 11 variables. The Original Investigators had recorded total years of smoking for current- and former-smokers directly from the questionnaire responses. The participants' answer to "total years of smoking" did not always match the number of years calculated from their responses to "age began smoking" and "age quit smoking". Other coding conventions limited the hours per day of passive exposure to smoke.

The Audit Team found good consistency between the Analytic.files and the questionnaires. We found no inconsistencies for five of the eight smoking variables; the other three had one error per variable. Likewise, we found no inconsistencies in two of the variables regarding passive

exposure to smoke, but found two inconsistencies for exposure to "passive smoke elsewhere".

Alcohol Intake Three variables provided information on intake of alcohol. We found no errors in these data.

Occupational Exposures Six variables were used to record occupational exposures to asbestos, chemicals/acids/solvents, coal/stone dust, coal tar/pitch/asphalt, diesel engine exhaust, and formaldehyde. These variables had then been collapsed into one variable for the statistical analyses by identifying a participant as occupationally exposed if a "1" for "yes" appeared for any one of the six variables. The Audit Team detected no errors in any of these six variables.

Education Although the education variable presented gradations in years of education, the final analyses compared those with and those without a high school education. We detected no errors in this latter variable.

Vital Status and Date of Death for Deceased Members of the Questionnaire Subset We identified 11 subjects from the questionnaire subset who were deceased. The date of death was represented by the month and year of death. We verified the vital status for members of the subset using two sources of information: a list of participants who took part in an American Cancer Society Nutrition Survey (conducted after completion of the ACS Study) and Social Security information available on the Internet.

However, to confirm that all individuals in the questionnaire subset identified as alive had indeed been alive at the ending date of the study, we needed additional information. At the request of the Audit Team, the ACS staff submitted the list of 250 names to the National Center for Health Statistics. There, technicians searched the NDI records for deaths that occurred during the study follow-up period of 1982–1989; they identified 242 records as possible matches for 71 individuals in the questionnaire subset. The Audit Team then reviewed each record, comparing the ACS and NDI entries for full name, SSN when available, date of birth (month and year), sex, race, marital status, state of residence, state of birth, and date of death (month and year).

By reviewing the NDI records, the Audit Team documented the month and year of death for the 11 individuals from this sample that had been identified as deceased by the ACS. For the other 60 individuals for whom one or more possible matches were detected, the Audit Team concluded that none of the possible NDI records represented subset members. Three cases were reviewed closely because they had no SSN in the ACS records and they matched NDI records on the basis of first and last name, birth month and year, sex, and race. However, the match was not consistent with the ACS records for state of residence or state of birth. Therefore, the Audit Team concluded that these possible matches did not reflect deaths of subset members and that the ACS coding of vital status was consistent with NDI records.

Thus, the Audit Team confirmed the vital status and dates of death (for 11 individuals) for all members of the questionnaire subset.

Survival Time The Audit Team recalculated this variable for each subject in the questionnaire subset. We noted no errors in the calculations and no inconsistences between this variable and the ascertainment of vital status previously described.

Subset of Deceased Subjects: Death Certificates

We audited the data given to the Reanalysis Team against source documents provided for a random sample of 250 deceased subjects.

We drew the random sample from deaths that had occurred during the first 6 years of the study, the original length of follow-up. The Original Investigators had added a seventh year of follow-up. Deaths only among men during this seventh year were included in the analysis. This oversight had been detected before the Atlanta audit and Ms Lally had completed the follow-up of women and had redone the analysis.

The Original Investigators provided the Audit Team with a listing of full names and dates of birth for the 250 subjects in the subset of deceased subjects. Of the 250 death certificates we requested, the Original Investigators retrieved 240 that had legible cause-of-death information. The ten missing or incomplete death certificates included six with missing microfiche roll and file information, one identified as "destroyed", one microfiche record that was blurred, one with an illegible cause of death, and one missing the cause-of-death section. For the 242 deaths we could verify, all had occurred before December 31, 1989, the study's ending date.

We audited the following variables by comparing information in the ACS Analysis.files to the death certificate copies:

- date of birth in Analysis.files versus date of birth on the death certificate;
- date of death in Analysis.files versus the date of death on the retrieved death certificate;

- subject identification information (SID, full name, birth date) in Analysis.files versus the same information on the death certificate; and
- ICD-9 cause-of-death code in Analysis.files versus code interpreted by the Audit Team nosologist from the death certificate.

Date of Birth The Audit Team found 11 dates of birth on death certificates that did not match the dates of birth in the Analysis.files, which had been derived from the participant's own entry on the questionnaire.

Date of Death We noted two inconsistencies, one in the month and one in the year of death.

Correct Death Certificate Due to variations in spelling of last names, or differences in dates of birth, or both, the Audit Team could not verify 15 death certificates as pertaining to the study subjects identified. We forwarded the SID numbers for these individuals to Dr Calle, who returned addresses, states of birth, names of spouses, and SSNs when available. We were then able to verify that all but one death certificate represented the appropriate study participant. That one death certificate clearly did not represent the intended study subject because the match had been based only on the phonetic spelling of the last name and the state of death. The Audit Team tracked the actual subject using Social Security information available on the Internet.

Cause-of-Death Codes As described for the Six Cities Study, the Audit Team nosologist compared the cause-of-death code with the one in the Analysis.files.

The variable containing cause-of-death information included either a two-digit CPS-II code (code book provided by Dr Calle) or a four-digit ICD-9 code. The two-digit code was a consolidation of ICD-9 codes. If a two-digit entry appeared, the Audit Team nosologist converted her ICD-9 code to the broader two-digit code on the basis of entries in the code book and then compared her code with that in the Analysis.files.

In 15 (6.3%) of the 240 death certificates with legible cause-of-death information, the Audit Team's two- or four-digit code did not match the code in the Analysis.files. Broad disease categories for cause-of-death analyses had been used by the Original Investigators. In 4 (1.6%) of the 240 death certificates, using the Audit Team's code would have altered the broad disease category. Details of these and other discrepancies are shown in Table 23.

The Audit Team next tracked how information had been incorporated into the broad disease categories used in the original analyses. Ms Lally again provided programming documentation. The program identified the cause-of-death code as codetype=1 (the two-digit CPS-II code) or codetype=2 (the four-digit ICD-9 code) and then proceeded with the following algorithm:

- asthma deaths were identified if code1 = 16 or 4930–4939;
- cardiopulmonary deaths were identified if code1 = 01, 03, 04, 05, 06, 07, 13, 14, 15, 16, 17, 18, 4010–4059, 4100–4179, 4200–4389, 4400–4409, 4800–4969, 4600–4789, or 5000–5199; due to the "else if" command used in each section, asthma deaths would not be included in this category because they had already been identified in an earlier step as belonging to the category of asthma deaths;
- lung cancer deaths were identified if code1 = 62 or 1620–1629; and
- all deaths not belonging to the first three groups were classified as "other".

The Audit Team detected a minor error in the computer program: the two-digit codes of 0A and 0B were coded as "other". However, as 0A referred to ICD-9 code 416.9, chronic pulmonary heart disease unspecified, and 0B referred to 440.9, generalized and unspecified atherosclerosis, these deaths should have been coded as cardiopulmonary, yet the program assigned them to the default "other" category. The Audit Team brought this to the attention of Dr Calle and Ms Lally, who searched the databases for individual records with a code 0A or 0B. For the total cohort, 16 deaths had been coded as 0A and 55 deaths had been coded as 0B. These 71 deaths had been grouped with "other" deaths rather than with cardiopulmonary deaths. The Audit Team concluded that this small number of additional cardiopulmonary deaths would not have affected the original results from the ACS Study.

AUDIT OF AIR QUALITY DATA

The ACS Study was not originally designed as an air pollution study. The air quality monitoring data used for the ACS analyses came from various sources, some of which are now technologically difficult to access. Documentation of the statistical reduction procedures has been lost. Summary statistics for different groups of standard metropolitan statistical areas had been derived by different investigators. These data sources do not indicate whether the tabulated values refer to all or a subset of monitors in a region or whether they represent means or medians. Values of sulfate for some cities could have come from several different sources. No information was available on any procedures that may have been applied to screen data. It

Table 23. Discrepancies in Codes Assigned to Causes of Death on Death Certificates Used by the ACS Study

Code in Analytic.files	Causes of Death on Death Certificate	Comments	Code by Audit Team's Nosologist
59 (159): Malignant neoplasm of other/ill-defined sites within digestive organs	Line a: Metastasis adenocarcinoma (primary unknown)	Adenocarcinoma does not necessarily originate in digestive organs (eg, lung adenocarcinoma)	99 (199.0): Malignant neoplasm without specification of site, disseminated
10: Thrombosis	Line a: Acute pulmonary embolism Line b: Thrombophlebitis, lower extremities Line c: Severe hypercalcemia with venous stasis	One other case also had pulmonary embolism (line a) and venous stasis (line b) on death certificate, yet had been coded as a 4 rather than as a 10 like this case	4 (415.1): Pulmonary embolism[a]
01: Ischemic heart disease	Line a: Septic shock Line b: Overwhelming septicemia		28 (038.9): Septicemia[a]
57: Cancer of pancreas	Line a: Cancer, liver with hepatic coma; pancreatitis		55 (155.2): Liver cancer
05: Other forms of heart disease (includes congestive heart failure)	Line a: Cardiopulmonary arrest Line b: Class IV congestive heart failure Line c: Renal failure		23 (586): Renal failure[a]
22: All other digestive diseases	Line a: Gangrene of large and small bowel Line b: Portal vein thrombosis Line c: Lactic acidosis		10 (453.8): Thrombosis of other specified veins
01: Ischemic heart disease	Line a: Bronchopneumonia	Atherosclerotic heart disease was listed in Part II (other significant conditions)	13: Pneumonia
414.0: Coronary atherosclerosis	Line a: Intracerebral hemorrhage (days) Line b: Atherosclerotic heart disease (years)		06 (431.0): Cardiovascular aneurysm (stroke)
73: Cancer of skin	Line a: Metastatic squamous cell carcinoma	Squamous cell carcinoma does not necessarily originate from skin (eg, lung squamous cell carcinoma)	99 (199.0): Malignant neoplasm without specification of site, disseminated
54: Cancer of rectum	Line a: Disseminated intravascular coagulopathy Line b: Colon cancer with liver metastasis		53: Cancer of colon

(Table continues next page)

[a] If the Audit Team's code were used, the grouping of diseases would have changed in the final analysis.

Table 23 (*continued*). Discrepancies in Codes Assigned to Causes of Death on Death Certificates Used by the ACS Study

Code in Analytic.files	Causes of Death on Death Certificate	Comments	Code by Audit Team's Nosologist
410.0: Acute mycocardial infarction	Line a: Cardiopulmonary arrest Line b: Acute myocardial infarction Line c: Atherosclerotic heart disease Line d: Acute myelogenous leukemia		36 (205.0): Acute myelogenous leukemia (leukemia)[a]
53: Colon cancer	Line a: Adenocarcinoma, abdomen, generalized	Adenocarcinoma in abdomen of woman is not necessarily colon cancer, could also be endometrial (uterus) or other parts of digestive tract	59 (159): Malignant neoplasm within digestive organs and peritoneum
05: Heart disease	"Deferred", then in different writing, the notation "4292"; "Pending" was written in the block for "Suicide, homicide, undetermined or pending investigation"	This would be the correct group for 429.2, cardiovascular disease, unspecified	Could not code, because a cause of death had not been determined
01: Ischemic heart disease	Line a: Cardiorespiratory arrest Line b: Arteriosclerotic heart disease Line c: Cardiovascular aneurysm		06: Cardiovascular aneurysm (stroke)
03: Hypertension	Line a: Congestive heart failure Line b: ACVD		05: Other forms of heart disease, includes congestive heart failure
20: Cirrhosis of liver	Line a: Sepsis Line b: Intestinal infarction	Part II (other significant conditions) noted alcoholic cirrhosis, but this section is not coded as the underlying cause of death	22 (557.0): Vascular insufficiency of intestine

[a] If the Audit Team's code were used, the grouping of diseases would have changed in the final analysis.

was not possible to audit instrument operating logs, filter weights, or other raw data records.

VALIDATION OF THE ORIGINAL ACS STUDY ANALYSES

The Reanalysis Team completed the validation using the SAS datasets provided by the Original Investigators. We used the same variables, the same criteria, and the same methods to replicate the results reported by Pope and colleagues (1995).

We estimated the mortality risk ratios with multiple regression analyses using the Cox proportional-hazards regression model (Cox 1972) as implemented in the SAS program. We computed mortality risk ratios (and their associated 95% CIs) due to sulfate and fine particle air pollution for lung cancer, cardiopulmonary disease, and all-cause mortality. As in the original analyses, we controlled for smoking, education, BMI, and other risk factors. We stratified all analyses by 5-year age categories, gender, and race (white, black, and other) and calculated separate baseline hazards for each age-sex-race stratum.

The complete results of the reanalysis are included in two appendices, which are available from the Health Effects Institute upon request: Appendix G. Computer Programs Used in the Replication of the American Cancer Society Study, and Appendix H. Replication of the Original Analysis of the American Cancer Society Study (Based on the Subcohort Used by the Original Investigators).

Validation of the Cohort Selection Process

The CPS-II cohort included 1,185,102 participants. Because only a subset of that cohort was used in the ACS Study, the Reanalysis Team first replicated the selection process. We selected all participants who lived within each MA for which data on sulfate or fine particle pollutants were available. To do this, we used a program that mapped the participants' ZIP codes onto MAs (see Part II for a further discussion of these methods). This procedure resulted in two population subcohorts, those used for the sulfate analyses (referred to as the sulfate cohort) and those used for the fine particle analyses (referred to as the fine particle cohort). Next, we excluded those participants for whom relevant information was missing. Using these two procedures, the Reanlysis Team selected 559,049 individuals for the sulfate cohort and 298,817 individuals for the fine particle cohort. Because a number of the MAs had pollution data regarding both fine particles and sulfate, some participants were members of both cohorts.

We found a different number of subjects than had been reported by the Original Investigators: 552,138 individuals in the sulfate cohort and 295,223 individuals in the fine particles cohort. Thus, the Reanalysis Team assigned 6,911 more subjects to the sulfate cohort and 3,594 more individuals to the fine particle cohort. The Original Investigators confirmed that this discrepancy was due to a typographic error in coding the formula used to determine the number of years that female former-smokers had been free of smoking. Consequently, the original SAS program had assigned a "missing" value to this variable and mistakenly excluded these individuals (7,706 female former-smokers in total).

When we began the reanalysis, the Original Investigators pointed out two other oversights in the original analyses. First, whereas the original publication had reported that deaths had been determined until December 31, 1989, only women who died before September 1, 1988, were included, thus excluding 5,421 female deaths. Second, they had intended that deaths from asthma would be categorized with deaths from cardiopulmonary disease. Instead, a computing error included these subjects in the all-cause mortality group. Because of this error, 83 asthma deaths (in men and women) had been coded incorrectly.

Results of the Reanalysis

For the first part of the validation analysis, we used the same cohort that the Original Investigators had used. For the second part, we included the 7,706 female former-smokers and the 5,421 female deaths that had been inadvertently left out of the original analyses. We also treated the 83 asthma deaths as cardiopulmonary deaths in this analysis.

Characteristics of Subjects in the ACS Analytic Cohort and Air Pollution Levels The Reanalysis Team assessed the following characteristics of the study population and the air pollution indices: number of MAs for each pollutant index, number of subjects, number of deaths, mean age at enrollment, sex, race, a profile of subjects' smoking experiences (cigarettes/day, number of years smoked, pipe/cigar smoker, and passive exposure to smoke), occupational exposure, education level, BMI, alcohol use, and exposure to air pollutants.

To compare the Original Investigators' results with those of the Reanalysis Team, Table 24 provides summary profiles of the original analytic cohort derived from CPS-II and the two indices of exposure to particulate air pollution: mean concentrations of sulfate particles for 1980 in the participants' areas of residence (derived from the US EPA's AIRS database) and median fine particle concentrations for 1979 through 1983 (calculated from the EPA's dichotomous sampler network). (The original results were presented in Table 1 of the AJRCCM publication.)

Although we confirmed the mean concentration of sulfate particles to be 11.0 $\mu g/m^3$, we calculated the SD to be 3.3 $\mu g/m^3$ rather than 3.6 $\mu g/m^3$. We also found the SD for fine particles to be 4.4 $\mu g/m^3$, slightly lower than the Original Investigators' value of 5.1 $\mu g/m^3$.

In the second part of the validation, which included the 7,706 female former-smokers and the additional 5,421 female deaths, we calculated a total of 43,361 deaths in the revised sulfate cohort of 559,049 individuals and 23,093 deaths among 298,817 individuals in the revised fine particle cohort. The percentage of females increased from 56.0% to 56.6% in the sulfate cohort and from 55.9% to 56.4% in the fine particle cohort. The percentages of current-smokers decreased slightly in both cohorts, whereas the percentage of former-smokers increased slightly. We also noted some small differences in the duration and intensity of smoking among former-smokers. The percentage of individuals subject to occupational exposures decreased slightly in both cohorts.

Table 24. Summary Characteristics of Subjects in the ACS Study's Analytic Cohort and in the Reanalysis Cohort

Characteristic	Original Analysis[a]		Validation Reanalysis[b]	
	Analysis with Sulfate Particles	Analysis with Fine Particles	Analysis with Sulfate Particles	Analysis with Fine Particles
Number of metropolitan areas	151	50	151	50
Number of subjects	552,138	295,223	***559,049***	***298,817***
Number of deaths	38,963	20,765	***43,361***	***23,093***
Age at enrollment (mean)	56.5	56.6	***56.6***	56.6
Sex (% female)	56.0	55.9	***56.6***	***56.4***
Race (%)				
White	94.2	94.0	94.2	94.0
Black	4.1	4.1	4.1	4.1
Other	1.7	1.9	1.7	1.9
Current cigarette smokers (%)	22.0	21.6	***21.7***	***21.4***
Cigarettes/day (mean)	22.0	22.1	22.0	22.1
Years smoked (mean)	33.5	33.5	33.5	33.5
Former cigarette smokers (%)	29.1	29.4	***30.0***	***30.2***
Cigarettes/day (mean)	22.0	22.0	***21.5***	***21.6***
Years smoked (mean)	22.3	22.2	***22.2***	***22.0***
Pipe/cigar smokers only (%)	4.1	3.9	***4.0***	3.9
Passive smoke (mean hours/day)	3.2	3.2	3.2	3.2
Occupational exposure (%)	20.0	19.5	***19.8***	***19.4***
Less than high school education (%)	12.3	11.3	12.3	11.3
Body mass index (mean)	25.1	25.0	25.1	25
Alcohol (mean drinks/day)	1.0	1.0	1.0	1
Sulfate particles ($\mu g/m^3$)				
Mean	11.0		11.0	
SD	3.6		***3.3***	
Range	3.6–23.5		3.6–23.5	
Fine particles ($\mu g/m^3$)				
Average median		18.2		18.2
SD		5.1		***4.4***
Range		9.0–33.5		9.0–***33.4***

[a] Original results from Pope et al 1995; corresponds to Table 1 in the original publication (reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine,* the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association).

[b] Reanalysis results based on revised ACS cohort. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

Adjusted Mortality Risk Ratios by Cause of Death for Cigarette Smoking and for a Difference in Pollution

Mortality risk ratios were calculated by replacing the variable representing the city (MA) in the statistical model with a continuous, linear variable representing either the mean of ambient sulfate or the median of fine particles. In this way, an exposure-response pattern was estimated according to level of pollution. Following the Original Investigators, we expressed the mortality risk ratios for an increase in particles across their entire ranges (Table 24). For sulfate particles, this factor was 19.9 $\mu g/m^3$ and for fine particles it was 24.5 $\mu g/m^3$.

The relative risk of mortality among current-smokers was derived by multiplying the relative risks associated with a series of smoking variables. These variables included indicators for current smoking status, daily consumption of cigarettes, and number of pack-years. In practice, this summary measure of risk was calculated by taking the exponential of the sum of the logarithm of the individual risks associated with these variables. The risk of mortality calculated in this manner assumed that, on average, a current-smoker consumed 20 cigarettes a day and had 25 pack-years at enrollment compared with a never-smoker.

Table 25a. Adjusted Mortality Risk Ratios by Cause of Death for Cigarette Smoking and for a Difference in Pollution: Original Result of the ACS Study[a]

Cause of Death	Current-Smoker[b]	Sulfate[c] (19.9 μg/m^3)	Fine Particles[c] (24.5 μg/m^3)
All	2.07 (1.75–2.43)	1.15 (1.09–1.22)	1.17 (1.09–1.26)
Lung cancer	9.73 (5.96–15.9)	1.36 (1.11–1.66)	1.03 (0.80–1.33)
Cardiopulmonary disease	2.28 (1.79–2.91)	1.26 (1.16–1.37)	1.31 (1.17–1.46)
All other	1.54 (1.19–1.99)	1.01 (0.92–1.11)	1.07 (0.92–1.24)

[a] From Pope et al 1995; corresponds to Table 2 in the original publication (reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine,* the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association). The difference in pollution equals the most-polluted areas compared with the least-polluted areas using either the sulfate or fine particle concentration as the measure of combustion-source air pollution. Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure.

[b] Risk ratios for cigarette smoking are estimated from the model using sulfate data and correspond to the risk of death for a current-smoker with 25 years of smoking 20 cigarettes per day compared with a never-smoker.

[c] Risk ratios have also been adjusted for cigarette smoking.

Table 25b. Adjusted Mortality Risk Ratios by Cause of Death for Cigarette Smoking and for a Difference in Pollution: Renalysis Results for the ACS Study Using the Same Analytic Cohort[a]

Cause of Death	Current-Smoker[b]	Sulfate[c] (19.9 μg/m^3)	Fine Particles[c] (24.5 μg/m^3)
All	2.07 ***(1.98–2.16)***	1.15 (1.09–1.22)	1.17 (1.09–1.26)
Lung cancer	9.73 ***(8.31–11.39)***	1.36 (1.11–1.66)	1.03 (0.80–1.33)
Cardiopulmonary disease	2.28 ***(2.14–2.43)***	1.26 (1.16–1.37)	1.31 (1.17–1.46)
All other	1.54 ***(1.44–1.64)***	1.01 ***(0.93***–1.11**)**	1.07 ***(0.95–1.21)***

[a] The difference in pollution equals the most-polluted areas compared with the least-polluted areas using either the sulfate or fine particle concentration as the measure of combustion-source air pollution. Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, aducation, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] Risk ratios for cigarette smoking are estimated from the model using sulfate data and correspond to the risk of death for a current-smoker with 25 years of smoking 20 cigarettes per day compared with a never-smoker.

[c] Risk ratios have also been adjusted for cigarette smoking.

Table 25c. Adjusted Mortality Risk Ratios by Cause of Death for Cigarette Smoking and for a Difference in Pollution: Reanalysis Results for the ACS Study Based on the Revised ACS Cohort[a]

Cause of Death	Current-Smoker[b]	Sulfate[c] (19.9 μg/m^3)	Fine Particles[c] (24.5 μg/m^3)
All	***2.06*** **(*1.97–2.14*)**	***1.16*** **(*1.10–1.23*)**	***1.18*** **(*1.10–1.27*)**
Lung cancer	***10.13*** **(*8.73–11.76*)**	1.36 **(*1.13–1.65*)**	***1.02*** **(**0.80–***1.30*)**
Cardiopulmonary disease	***2.31*** **(*2.17–2.46*)**	***1.28*** **(*1.19–1.39*)**	***1.32*** **(*1.19***–1.46**)**
All other	***1.50*** **(*1.41–1.60*)**	***1.02*** **(*0.93***–1.11**)**	***1.09*** **(*0.98–1.22*)**

[a] The difference in pollution equals the most-polluted areas compared with the least-polluted areas using either the sulfate or fine particle concentration as the measure of combustion-source air pollution. Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, aducation, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

[b] Risk ratios for cigarette smoking are estimated from the model using sulfate data and correspond to the risk of death for a current-smoker with 25 years of smoking 20 cigarettes per day compared with a never-smoker.

[c] Risk ratios have also been adjusted for cigarette smoking.

Tables 25a, 25b, and 25c present adjusted mortality risk ratios (and 95% CIs) by cause of death for current-smokers and for an increase of 19.9 $\mu g/m^3$ sulfate or an increase of 24.5 $\mu g/m^3$ fine particles (Table 25a gives original results as they were presented in Table 2 of the original publication; Table 25b shows the Reanalysis Team's results using the same cohort; and Table 25c shows results using the revised cohort). The mortality risk ratios were adjusted for age, sex, race, cigarette smoking, passive exposure to cigarette smoke, BMI, drinks per day of alcohol, education, and occupational exposure.

The Original Investigators used a conservative method of calculating CIs on the mortality risk ratios for current-smokers. (For a complete description of the different formulas to calculate CIs used by the Original Investigators and by the Reanalysis Team, see the Six Cities Study section Adjusted Mortality Rate Ratios for Current-Smokers, Former-Smokers, and for the Most-Polluted Versus the Least-Polluted City by Cause of Death.) Using the method preferred by the Reanalysis Team, the CIs for current-smokers were somewhat narrower (see Table 25c) than those calculated by the Original Investigators. For example, the original 95% CI of 5.96–15.9 for lung cancer mortality among current-smokers decreased in width to 95% CI: 8.31–11.4.

When we included additional data in the second part of the validation analysis, the mortality risk ratios for both sulfate and fine particle exposure tended to increase. For example, the mortality risk ratio for deaths from cardiopulmonary disease associated with sulfate exposure increased from 1.26 (95% CI: 1.16–1.37) to 1.28 (95% CI: 1.19–1.39).

Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status Tables 26 and 27 summarize the adjusted mortality risk ratios by gender and smoking status for an increase in particles across their ranges for deaths due to all causes, lung cancer, and cardiopulmonary disease. The mortality risk ratios were adjusted for age, sex, race, cigarette smoking, passive exposure to cigarette smoke, BMI, drinks per day of alcohol, education, and occupational exposures. (Tables 26a and 27a give original results as they were presented in Table 3 of the AJRCCM publication; Tables 26b and 27b show the Reanalysis Team's results using the same cohort; and Tables 26c and 27c show the Reanalysis Team's results using the revised cohort.)

Although the first part of the validation analysis produced only trivial discrepancies between the Reanalysis Team's results and those of the Original Investigators (Tables 26b and 27b), including additional data in the second part of the validation analyses again tended to increase the estimates of the mortality risk ratios. For example, the mortality risk ratios for female ever-smokers increased in three analyses: (1) for all causes of death associated with sulfate exposure (Tables 26a and 26c), it increased from 1.14 (95% CI: 0.97–1.33) to 1.18 (95% CI: 1.04–1.35); (2) for cardiopulmonary deaths associated with sulfate exposure it increased from 1.30 (95% CI: 1.01–1.66) to 1.44 (95% CI: 1.17–1.78); and (3) for cardiopulmonary deaths associated with fine particle exposure (Tables 27a and 27c), it increased from 1.27 (95% CI: 0.92–1.74) to 1.32 (95% CI: 1.01–1.72).

Table 26a. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Sulfate (19.9 $\mu g/m^3$): Original Results of the ACS Study[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	1.15 (1.09–1.22)	1.36 (1.11–1.66)	1.26 (1.16–1.37)
Women	1.18 (1.06–1.30)	1.17 (0.80–1.72)	1.39 (1.20–1.61)
Men	1.14 (1.06–1.23)	1.43 (1.13–1.81)	1.20 (1.08–1.33)
Never-smokers	1.18 (1.06–1.30)	1.51 (0.73–3.11)	1.36 (1.19–1.58)
Women	1.20 (1.06–1.36)	1.61 (0.66–3.92)	1.44 (1.20–1.74)
Men	1.14 (0.97–1.34)	1.36 (0.40–4.66)	1.28 (1.03–1.58)
Ever-smokers	1.14 (1.06–1.23)	1.35 (1.10–1.66)	1.20 (1.08–1.33)
Women	1.14 (0.97–1.33)	1.10 (0.72–1.68)	1.30 (1.01–1.66)
Men	1.14 (1.05–1.24)	1.44 (1.14–1.83)	1.17 (1.05–1.32)

[a] From Pope et al 1995; corresponds to the left half of Table 3 in the original publication (reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine,* the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association). Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure.

Table 26b. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Sulfate (19.9 μg/m^3): Renalysis Results for the ACS Study Using the Same Analytic Cohort[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	1.15 (1.09–1.22)	1.36 (1.11–1.66)	1.26 (1.16–1.37)
Women	1.18 (***1.07***–1.30)	1.17 (0.80–1.72)	1.39 (1.20–***1.62***)
Men	1.14 (***1.05–1.22***)	1.43 (1.13–1.81)	1.20 (1.08–1.33)
Never-smokers	1.18 (***1.07***–1.30)	1.51 (0.73–3.11)	***1.38*** (***1.20***–1.58)
Women	1.20 (1.06–***1.37***)	1.61 (0.66–3.92)	***1.45*** (***1.21–1.75***)
Men	1.14 (0.97–1.34)	1.36 (0.40–4.66)	1.28 (1.03–1.58)
Ever-smokers	1.14 (1.06–1.23)	1.35 (1.10–1.66)	1.20 (1.08–1.33)
Women	***1.15*** (***0.98–1.34***)	1.10 (0.72–1.68)	1.30 (1.01–1.66)
Men	1.14 (1.05–***1.23***)	1.44 (***1.13***–1.83)	1.17 (1.05–1.32)

[a] Values are risk ratios (95% CIs). Risk ratios have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

Table 26c. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Sulfate (19.9 μg/m^3): Reanalysis Results for the ACS Study Based on the Revised ACS Cohort[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	***1.16*** (***1.10–1.23***)	1.36 (***1.13–1.65***)	***1.28*** (***1.19–1.40***)
Women	***1.20*** (***1.10***–1.30)	***1.22*** (***0.87–1.70***)	***1.42*** (***1.25–1.62***)
Men	1.14 (***1.05–1.22***)	1.43 (1.13–1.81)	1.20 (1.08–***1.31***)
Never-smokers	***1.19*** (***1.08***–1.30)	***1.87*** (***0.95–3.69***)	***1.37*** (***1.20–1.56***)
Women	***1.21*** (***1.08***–1.36)	***2.17*** (***0.96–4.88***)	***1.42*** (1.20–***1.67***)
Men	1.14 (0.97–1.34)	1.36 (0.40–4.66)	1.28 (1.03–1.58)
Ever-smokers	***1.15*** (***1.07–1.24***)	***1.33*** (***1.09–1.62***)	***1.23*** (***1.12–1.34***)
Women	***1.18*** (***1.04–1.35***)	***1.09*** (***0.75–1.57***)	***1.44*** (***1.17–1.78***)
Men	1.14 (1.05–***1.23***)	1.44 (***1.13***–1.83)	1.17 (1.05–1.32)

[a] Values are risk ratios (95% CIs) which have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

Table 27a. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Fine Particles (24.5 μg/m^3): Original Results of the ACS Study[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	1.17 (1.09–1.26)	1.03 (0.80–1.33)	1.31 (1.17–1.46)
Women	1.16 (1.02–1.32)	0.90 (0.56–1.44)	1.45 (1.20–1.78)
Men	1.18 (1.07–1.30)	1.10 (0.81–1.47)	1.24 (1.08–1.41)
Never-smokers	1.22 (1.07–1.39)	0.59 (0.23–1.52)	1.43 (1.18–1.72)
Women	1.21 (1.02–1.43)	0.65 (0.21–2.06)	1.57 (1.23–2.01)
Men	1.24 (1.00–1.54)	0.49 (0.09–2.66)	1.24 (0.93–1.67)
Ever-smokers	1.15 (1.05–1.26)	1.07 (0.82–1.39)	1.24 (1.08–1.42)
Women	1.10 (0.90–1.33)	0.95 (0.57–1.58)	1.27 (0.92–1.74)
Men	1.16 (1.05–1.29)	1.12 (0.83–1.52)	1.23 (1.06–1.43)

[a] From Pope et al 1995; corresponds to the right half of Table 3 in the original publication (reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine,* the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association). Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure.

The Original Investigators reported that lung cancer mortality had not been associated with combustion-source air pollution when fine particles (in 50 MAs) were used as the pollution index; however, they had found an association when sulfate (in 151 MAs) was used as the index. The Original Investigators had considered whether the difference in MAs might account for the different findings. To test this hypothesis, they had restricted their analyses to the 47 MAs for which both sulfate and fine particle data were available. Again, no association was found when fine particles were used as the pollution index. However, when sulfate was used as the index, the adjusted mortality risk ratio for lung cancer was 1.20 (95% CI: 1.08–1.34) and for cardiopulmonary disease it was 1.44 (95% CI: 1.11–1.86). Using the same dataset as the Original Investigators, the Reanalysis Team reproduced these results.

The Original Investigators had also reported that high, low, and mean temperatures were not correlated with either sulfate or fine particle pollution. However, they had found that sulfate particle levels were slightly lower in both relatively cold (normal mean temperatures lower than 50°F) and relatively hot (normal mean temperatures higher than 60°F) MAs. When these weather indicator variables were included in the risk models, the adjusted mortality risk

Table 27b. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Fine Particles (24.5 μg/m^3): Reanalysis Results for the ACS Study Using the Same Analytic Cohort[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	1.17 (1.09–1.26)	1.03 (0.80–1.33)	1.31 (1.17–1.46)
Women	1.16 (1.02–1.32)	0.90 (0.56–1.44)	***1.46*** (1.20–***1.77***)
Men	1.18 (1.07–***1.29***)	1.10 (0.81–***1.48***)	1.24 (1.08–1.41)
Never-smokers	1.22 (1.07–1.39)	0.59 (0.23–1.52)	1.43 (1.18–1.72)
Women	1.21 (1.02–***1.44***)	0.65 (0.21–2.06)	***1.58*** (1.23–***2.02***)
Men	1.24 (1.00–1.54)	0.49 (0.09–2.66)	1.24 (0.93–***1.66***)
Ever-smokers	1.15 (1.05–1.26)	1.07 (0.82–1.39)	1.24 (1.08–1.42)
Women	1.10 (0.90–1.33)	0.95 (0.57–1.58)	1.27 (0.92–1.74)
Men	1.16 (1.05–1.29)	1.12 (0.83–1.52)	1.23 (1.06–1.43)

[a] Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

Table 27c. Adjusted Mortality Risk Ratios for the Most-Polluted Areas Compared with the Least-Polluted Areas for Deaths from All Causes, Lung Cancer, and Cardiopulmonary Disease by Gender and Smoking Status for Fine Particles (24.5 μg/m^3): Reanalysis Results for the ACS Study Based on the Revised ACS Cohort[a]

Study Group	All Causes	Lung Cancer	Cardiopulmonary Disease
All combined	***1.18 (1.10–1.27)***	***1.02*** (0.80–***1.30***)	***1.32 (1.19***–1.46)
Women	***1.19 (1.06–1.33)***	***0.89 (0.59–1.34)***	1.45 (***1.22–1.71***)
Men	1.18 (1.07–***1.29***)	1.10 (0.81–***1.48***)	1.24 (1.08–1.41)
Never-smokers	***1.24 (1.10–1.40)***	***0.73 (0.30–1.80)***	1.43 (***1.20–1.70***)
Women	***1.25 (1.07–1.45)***	***0.87 (0.30–2.52)***	***1.54 (1.24–1.92)***
Men	1.24 (1.00–1.54)	0.49 (0.09–2.66)	1.24 (0.93–***1.66)***
Ever-smokers	1.15 (1.05–1.26)	***1.04 (0.81–1.34)***	***1.25 (1.10***–1.42)
Women	***1.12 (0.95–1.32)***	***0.88 (0.56–1.39)***	***1.32 (1.01***–***1.72***)
Men	1.16 (1.05–1.29)	1.12 (0.83–1.52)	1.23 (1.06–1.43)

[a] Values are risk ratios (95% CIs), which have been adjusted for age, sex, race, cigarette smoking, exposure to passive cigarette smoke, body mass index, drinks per day of alcohol, education, and occupational exposure. Values in bold italics show Reanalysis Team results that differ from those reported by the Original Investigators.

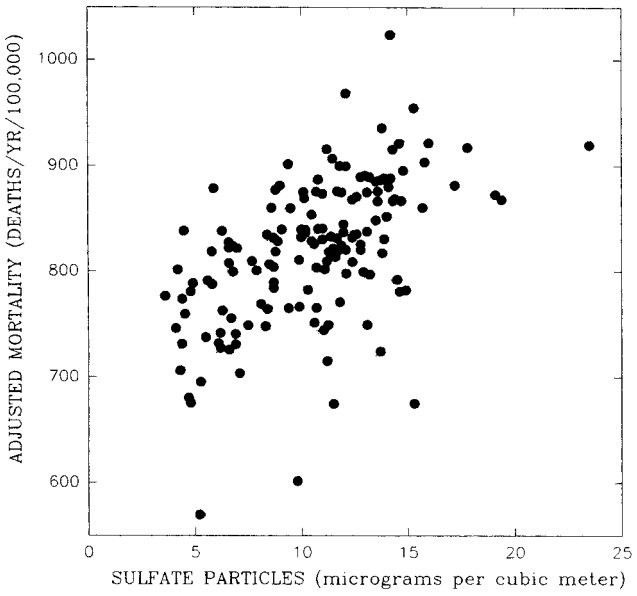

Figure 12. Age-, sex-, and race-adjusted population-based mortality rates for 1980 plotted against mean sulfate air pollution levels for 1980. Data are from metropolitan areas that correspond approximately to areas used in the prospective cohort analysis. Original results from Pope and colleagues 1995 (Figure 1; reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine*, the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association).

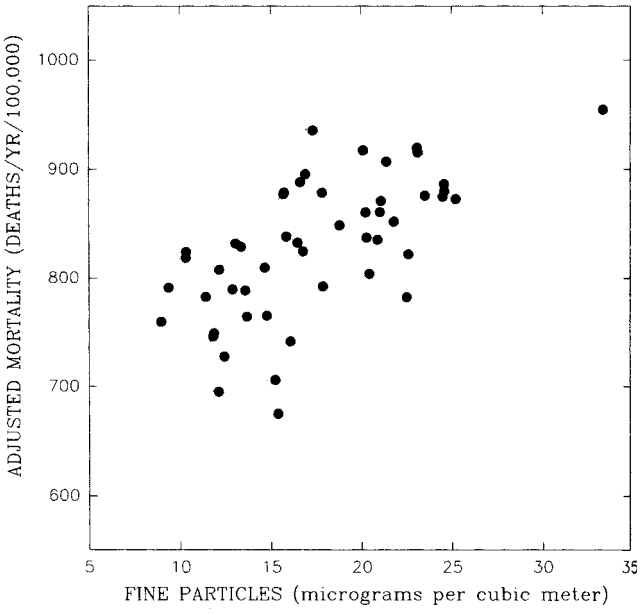

Figure 13. Age-, sex-, and race-adjusted population-based mortality rates for 1980 plotted against median [*not* mean, *as in the original publication*] **fine particulate air pollution levels for 1979 to 1983.** Data are from metropolitan areas that correspond approximately to areas used in the prospective cohort analysis. Original results from Pope and colleagues 1995 (Figure 2; reprinted with permission from the *American Journal of Respiratory and Critical Care Medicine*, the Official Journal of the American Thoracic Society; Copyright © 1995 American Lung Association).

ratios for deaths from lung cancer and cardiopulmonary disease were 1.36 (95% CI: 1.11–1.66) and 1.23 (95% CI: 1.13–1.34), respectively, when sulfate was used as the pollution index, and 1.05 (95% CI: 0.82–1.36) and 1.26 (95% CI: 1.13–1.40), respectively, when fine particles were used as the index. Aside from a minor difference in the risk ratio for cardiopulmonary mortality of 1.24 (95% CI: 1.14–1.35), the Reanalysis Team reproduced these results using the same dataset as the Original Investigators.

When asthma deaths were included in cardiopulmonary deaths, the risk ratio from sulfate changed marginally from 1.26 to 1.25 (95% CI: 1.15–1.36) and that from fine particles changed from 1.31 to 1.30 (95% CI: 1.16–1.44).

We also replicated their figures (Figures 1 and 2 in the AJRCCM publication) using population-based mortality rates for 1980 (adjusted for age, sex, and race) that were provided by the Original Investigators and found no discrepancies (Figures 12 and 13).

SUMMARY AND CONCLUSIONS

In Part I of the reanalysis of the Six Cities and ACS studies, we used two methods to ensure the validity of the original studies' results: a data quality audit and a series of validation analyses. As might be expected in studies as large and broad as these, we found some small discrepancies during the reanalysis. These discrepencies do not alter in any substantive fashion the results of the original analyses; thus, the Reanalysis Team is satisfied that the objectives of this reanalysis have been satisfied and that the original results are indeed correct as published.

The validation analysis subsequently conducted on both studies using the same data and the same methods the Original Investigators had used resulted in nearly complete agreement with the original findings. Discrepancies in the Six Cities Study were minor.

- Some typographic errors were found in the summary table in the Six Cities Study as reported in the NEJM. However, because those values had not been used in subsequent calculations, they had no effect on the findings of the study.
- Most of the discrepancies noted by the Reanalysis Team pertained to the estimates of pollutant levels; some of these are likely due to subtle differences in the order of calculations performed or in the software used.
- The validation analysis of the key results from the Six Cities Study attained complete agreement with all of the point estimates of the various rate ratios calculated. The only discrepancy we found was a minor typographic error in reporting the number of pack-years smoked.
- The Reanalysis Team updated the Six Cities cohort to include the missing person-years of observation identified through the data quality audit. Adding 928 person-years of observation resulted in an increase of 14 deaths in the six cities. Using the same methods of analysis as had been applied to the original cohort led to mortality rate ratios associated with exposure to fine particulate matter that were higher than those reported. For example, the original relative risk of 1.26 for all-cause mortality increased to 1.28 after adjusting for early censoring of person-years. This adjustment increased the mortality rate ratio for cardiopulmonary disease from 1.37 to 1.43.

While reconstructing the ACS database to match the information used in the original analysis, ACS staff and the Reanalysis Team noted three errors in computer programming: asthma deaths had been excluded from the total cardiopulmonary deaths, a group of female former-smokers had been excluded from the subcohort, and female deaths had been censored earlier than they should have been.

The Reanalysis Team reproduced the ACS results when we used the same data as those used to derive the findings reported in the AJRCCM. However, when we included the group of female former-smokers and the female deaths and asthma deaths that had been excluded, several differences became apparent.

- The mortality risk ratios due to both sulfate and fine particle exposure increased slightly.
- The mortality risk ratios increased when we compared the most-polluted city with the least-polluted city.
- The mortality risk ratio due to sulfate exposure became significant for all causes of death for female ever-smokers.

One further discrepancy the Reanalysis Team noted in both studies was that the methods the Original Investigators used to calculate CIs for mortality risk estimates related to tobacco consumption were incorrect. These methods had not been used for mortality risk estimates for ambient air pollution in either of the two studies. The Reanalysis Team chose to use a direct method, which emphasized the dependence between the parameter estimates, to calculate CIs on risk estimates for the effect of tobacco consumption on mortality. The direct method noticeably narrowed the CIs for the mortality estimates for both studies.

Table 28. Errors in the Data Used in the Six Cities Study and the ACS Study Found by the Reanalysis Team

Finding	Magnitude of Potential Impact	Effects on Validity	How Addressed in Part I[a]
Six Cities: Early censorship of time-on-study for some participants in some cities	Loss of approximately 1% of person-years, effect of which was greatest in Portage and Topeka; no early censorship in Watertown	Minor	Recalculated
ACS: Exclusion of a group of female former-smokers who met selection criteria	7,706 Women excluded	Small increase in mortality risk ratios for sulfate and fine particles	Recalculated
ACS: Follow-up of women curtailed on 9/1/88 rather than the correct date of 12/31/89	5,421 Deaths among women excluded from analysis	Small increase in mortality risk ratios for sulfate and fine particles	Recalculated
ACS: Asthma deaths included in "other" category, rather than as cardiopulmonary deaths	83 Asthma deaths excluded from cardiopulmonary category	Minor	Recalculated
ACS: Computer programming error resulted in two ICD-9 codes reflecting cardiopulmonary diseases not included in cardiopulmonary category	71 Deaths added to "other" rather than cardiopulmonary category	Minor	Recalculated

[a] The Reanalysis Team recalculated relative risks on the basis of including the data that had been excluded by the Original Investigators.

Overall, the Audit Team found that both studies had been well conducted and well documented. The minor errors that we found in the data would not have materially impacted the data as published or the Original Investigators' conclusions. The variables used in the original publications were valid. The error rate we calculated for each variable in each study was less than 5% and not critical from an epidemiologic standpoint with regard to changing the estimates of relative risk. The audits of both studies uncovered some systematic errors (Table 28). However, the Reanalysis Team was able to reconstitute the cohorts using the information from the data audits and carry out detailed reanalyses that showed minor differences from original findings.

The Reanalysis Team analyzed most of the data twice using different statistical packages (S-PLUS and SAS) and obtained the same results. This indicates that the numerical results were not dependent upon the computer programs that were used to fit the Cox proportional-hazards regression models in the Original Investigations.

Although Part I of the reanalysis of these two important cohort mortality studies effectively confirms the numerical results reported by the Original Investigators, a final assessment of these two studies was conducted in Part II. Whereas Part I of the reanalysis was based on the same data and methods used by the Original Investigators, in Part II of the reanalysis we tested the robustness of these validated findings to different methods of analysis. We also included additional data not considered by the Original Investigators.

REFERENCES

Abbey DE, Nishino N, McDonnell WF, Burchette RJ, Knutsen SF, Lawrence Beeson W, Yang JX. 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. Am J Respir Crit Care Med 159:373–382.

Bates D, Hayes CG, Knowles M. 1985. Geographic variation in declining ischemic heart disease mortality in the United States, 1968–1978. Am J Epidemiol 122:657–672.

Briggs SLK, Thurston GD, Dumyahn TS, Evans JS, Spengler JD, Turner WA. 1982. The effectiveness of dichotomous sampler quality assurance procedures. APCA Annual Meeting, New Orleans LA.

Burnett RT, Jessiman B, Stieb D, Krewski D. 1997. Health effects of ambient particulate matter: A Canadian perspective. In: Health Effects of Particulate Matter in Ambient Air. Prague, Czech Republic. Ed Jaroslav J Vostal. April 23–25, 1997.

Cochran WG. 1977. Sampling Techniques, 3rd ed. John Wiley & Sons, New York NY.

Courtney WJ, Shaw RW, Dzubay TG. 1982. Precision and accuracy of a beta gauge for aerosol mass determinations. Environ Sci Technol 16:236–239.

Coutant RW. 1977. Effect of environmental variables on collection of atmospheric sulfate. Environ Sci Technol 11:873–878.

Cox DR. 1972. Regression models and life tables (with discussion). J R Stat Soc B34:187–200.

Cox DR, Oakes D. 1983. Analysis of Survival Data. Chapman and Hall, London, England.

Crosby PB. 1979. Quality is Free. McGraw-Hill, New York NY.

Dockery D. 1996. The air we breathe: 22 years of the Six Cities Study [Harvard School of Public Health October 4, 1996. Around the School, Web site]. Available at: http://www.hsph.harvard.edu/around/oct96/oct4/sixcity.html. Accessed March 22, 2000.

Dockery DW, Schwartz J, Spengler J. 1992. Air pollution and daily mortality: Associations with particulates and acid aerosols. Environ Res 59:362–373.

Dockery DW. 1993. Percentile curves for evaluations of repeated measures of lung function. Occup Med 8:323–338.

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE. 1993. An association between air pollution and mortality in six US cities. N Engl J Med 329:1753–1759.

Dzubay TG, Barbour RK. 1983. A method to improve the adhesion of aerosol particles on Teflon filters. J Air Pollut Control Assoc 33:692–695.

Eaton WC, Arey FK, McKee JL. 1982. Systems Audit 1982: Harvard School of Public Health Six City Study. Report RTI/2218/94. Research Triangle Institute, Research Triangle Park NC.

Ferris BG, Speizer FE, Spengler JD, Dockery D, Bishop YMN, Wolfson M, Humber C. 1979. Effects of sulfur oxides and respirable particles on human health: Methodology and demography of populations in study. Am Rev Respir Dis 120:767–779.

Gamble JF. 1998. $PM_{2.5}$ and mortality in long-term prospective cohort studies: Cause-effect or statistical associations? Environ Health Perspect 106:535–549.

Harvard School of Public Health Air Quality Group. 1982. Quality Assurance Manual for Air Quality Assessment (Internal Document). Harvard School of Public Health, Boston MA.

Health Effects Institute. 1995. Particulate Air Pollution and Daily Mortality: Replication and Validation of Selected Studies, The Phase I.A Report of the Particle Epidemiology Evaluation Project. Health Effects Institute, Cambridge MA.

Health Effects Institute. 1997. Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants, The Phase I.B Report of the Particle Epidemiology Evaluation Project. Health Effects Institute, Cambridge MA.

Hoover BK, Baldwin JK. 1984. Meeting the quality assurance challenges of the 1980's: Team auditing by toxicologists and QA professionals. J Am Coll Toxicol 3:129–139.

Hoover BK, Baldwin JK, Uelner AF, Whitmire CE, Davies CL, Bristol DW. 1986. Managing Conduct and Data Quality of Toxicology Studies. Princeton Scientific Publishers, Princeton NJ.

Hoover BK, Foliart DE, Calisti LJ, White WH. 1999. Quality assurance audits of the Harvard Six Cities and ACS Studies. The Fifteenth Health Effects Institute Annual Conference, May 9–12, La Jolla CA. Health Effects Institute, Cambridge MA.

Koenig JQ, Larson TV, Hanley QS, Rebolledo V, Dumler K. 1993. Pulmonary function changes in children associated with fine particulate matter. Environ Res 63:26–38.

Koenig JQ, Larson TV, Norris G, Claiborn C, Finn D, Zweidinger R, Lewtas J, Schwartz J. 1998. A fine particulate composition variability and exacerbation of asthma [Particulate Matter Research Activities Web site]. January 24, 2000. Available at: http://www.pmra.org/pmra/PM.nsf. Accessed March 22, 2000.

Koutrakis P, Wolfson JM, Spengler JD. 1988. An improved method for measuring aerosol strong acidity: Results from a nine-month study in St. Louis, Missouri, and Kingston, Tennessee. Atmos Environ 22:157–162.

Lave LB, Seskin EP. 1970. Air pollution and human health. Science 169:723–733.

Lee ET. 1992. Statistical Methods for Survival Data Analysis, 2nd ed. John Wiley & Sons, New York NY.

Lipfert FW. 1993. Community Air Pollution Mortality: Analysis of 1980 Data from US Metropolitan Areas, I, Particulate Air Pollution. BNL 48446-R. US Department of Energy, Brookhaven National Laboratory, Upton NY.

Lipfert FW, Malone RG, Daum ML, Mendell NR, Yang CC. 1988. A Statistical Study of the Macroepidemiology of Air Pollution and Total Mortality. BNL 52122. US Department of Energy, Brookhaven National Laboratory, Upton NY.

Pope CA, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669–674.

Schilling EG, Sommers DJ. 1988. Acceptance sampling. In: Juran's Quality Control Handbook, 4th ed (Juran JM, Gryna FM, eds). McGraw-Hill, New York NY.

Schwartz J. 1994. What are people dying of on high air pollution days? Environ Res 64:26–35.

Schwartz J, Dockery DW, Neas LM. 1996. Is daily mortality associated specifically with fine particles? J Air Waste Manage Assoc 46:927–939.

Siconolfi RM. 1986. Statistical approaches to auditing. In: Managing Conduct and Data Quality of Toxicology Studies (Hoover BK et al, eds). Princeton Scientific Publishers, Princeton NJ.

Spengler JD, Briggs SLK, Özkaynak H. 1986. Relationships Between TSP Measurements and Size-Fractionated Particle Mass Measurements in Six Cities Participating in the Harvard Air Pollution Health Study (Report to OAQPS under EPA Cooperative Agreement CR-812667-01-0). US Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park NC.

Spengler JD, Thurston GD. 1983. Mass and elemental composition of fine and coarse particles in six US cities. J Air Pollut Control Assoc 33:1162–1171.

Spirtas R, Levin HJ. 1971. Patterns and trends in levels of suspended particulate matter. J Air Pollut Control Assoc 21:329–333.

Steelman RL. 1986. Industrial management view of quality assurance for Food and Drug Administration regulated products. In: Managing Conduct and Data Quality of Toxicology Studies (Hoover BK et al, eds). Princeton Scientific Publishers, Princeton NJ.

Venables WN, Ripley BD. 1997. Modern Applied Statistics with S-PLUS, 2nd ed. Springer-Varlag, New York NY

Wang Y, Krewski D, Bartlett S, Zielinski JM. 1994. The Effect of Record Linkage Errors on Statistical Inference in Cohort Mortality Studies. Technical Report 245. Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Ontario, Canada.

World Health Organization. 1975. *International Classification of Diseases, Ninth Revision*, Vol 1. World Health Organization, Geneva, Switzerland.

APPENDIX A. Data Audit Standards, Goals, and Plan

The practice of team auditing, which follows published techniques (Hoover and Baldwin 1984), was selected for the Reanalysis Project due to the scope of work. It is a robust approach to auditing because it combines the resources and expertise of individuals with different qualifications so the final result is greater than any one individual with a single expertise could have accomplished working separately. Each team member was chosen for his or her expertise so the group was not limited to the employees of any one company. Preliminary results of the Reanalysis Project audit were presented at the HEI Annual Conference in La Jolla CA in May 1999 (Hoover et al 1999).

The Audit Team coordinator for the whole project was Ms Kristin Hoover. She has more than 25 years of experience in developing and managing audits of projects for organizations such as university research programs and commercial analytical chemistry laboratories on topics such as methods development, toxicology, and epidemiology. Her recent efforts have involved quality assurance monitoring and auditing for other HEI epidemiologic studies related to air pollution.

Donna Foliart, MD, MPH, has been a subcontracting consultant with Ms Hoover for previous HEI studies. Dr Foliart obtained her MD with honors from the University of California, San Francisco, in 1978, her MPH from University of California, Berkeley, in 1981, and is board certified in preventative medicine/occupational medicine (1984). She is currently with the Public Health Institute in Berkeley, where she is the principal investigator of a study of childhood leukemia.

Warren White, BSc (mathematics, California Institute of Technology), MS and PhD (mathematics, University of Wisconsin), is an expert in air pollution monitoring data. He has served in a variety of positions including OAS visiting professor at the Institute of Pure and Applied Mathematics,

National Research Council, Rio de Janeiro, Brazil; visiting professor at Brookhaven National Laboratory; and is currently at Washington University, St Louis MO. Dr White's contribution to air pollution research is well known.

Ms Linda Calisti, BSc (University of Pittsburgh, 1971) has over 20 years of experience in various types of audits such as analytical chemistry, toxicology, and human clinical research.

GOALS AND STANDARDS

The overall objective of this audit was to conduct an independent, rigorous, retrospective, and defensible assessment of the raw data from original source documents and electronic data files from these studies to support the efforts of the Reanalysis Team. In the book *Quality is Free*, Crosby (1979) defines "quality" as conformance to requirements or standards (discussed in Hoover et al 1986). For the purposes of this audit, the standards used are described below.

Standards for the Retrospective Audits

Standards established by the Health Effects Institute for this project were summarized in the following internal project documents:

- The Health Effects Institute Epidemiology Reanalysis Project: Project Description (March 25, 1998);
- Statement of Specifications: Epidemiology Reanalysis Project Data Quality Audit; and
- Health Effects Institute Procedures for Quality Assurance and Quality Control (February 20, 1997).

The investigators on the Reanalysis Team described the following types of documents for the Audit Team to review so we could assess the internal standards the Original Investigators had established for their own studies.

Protocols for Each Study The Audit Team found no formal study protocols for either study. Instead, we audited the data against printouts of the electronic files provided to the Reanalysis Team. In the original publications, we found information that would normally be contained in a study design protocol and we audited the studies according to the published information.

Internal Standard Operating Procedures Used in the Study The Audit Team found no procedural rules for either study that were formally identified as Standard Operating Procedures. Therefore, we used whatever documentation existed for each study that explained the rules for data collection, manipulation, and inclusion or exclusion for analyses. For the Six Cities Study, we obtained four notebooks of coding rules that included discussions of coding problems and associated corrective actions. We found no explanatory documentation of coding for the ACS Study. Therefore, the Audit Team determined the applied coding conventions by inference.

Existing Quality Assurance Audits The Audit Team examined both internal and external audits for the Six Cities Study, but none were available for the ACS Study. However, the remaining contact at the ACS made available some computer programming documentation; this person had identified several programming problems, which were discussed in the main report.

General Audit Plan (Applicable to Both Studies)

A detailed quality assurance plan was prepared before the audit and submitted to HEI in March of 1999. The Audit Team followed this plan for both studies with some minor exceptions related to availability of documentation or time constraints that could not have been foreseen when the plan was developed. Ms Hoover acted as the principal contact with HEI and was responsible for leading this audit program. Teams were used for each onsite audit. The Audit Team identified the following types of information as applicable for a statistically relevant subset of data.

Organization and Personnel We used discussions with study personnel and written records to determine how the study had been organized and who had been responsible for management of the data. In both studies, the analyses were restricted to selected subsets of the cohort. The Audit Team determined how the questionnaires, death certificates, and air pollution data had been filed and what resources would be available to assist in the retrieval of records. The Audit Team also determined what personnel were still available who had actually worked on each study or had the greatest knowledge of procedures.

Data Collection For the Six Cities Study, the Audit Team evaluated the documentation of data collection and procedural methods. We audited the data against the established coding conventions and rules, and followed any changes in coding. We examined (1) documentation of how any discrepancies in coding had been resolved; (2) field restrictions to determine how they had been utilized; and (3) the circumstances and documents about "missing" data to determine that each instance had been treated consistently.

For the ACS Study, none of these items were available and we could not perform such a thorough audit.

Computer Processing The Audit Team reviewed changes to computer files to determine that they had been implemented consistently and whether the requirements as detailed in the published report had been followed. We compared changes to hard copies against the computer files and vice versa. We examined criteria for data reduction to ensure that they had been followed consistently. When we found documentation of a discrepant data element, we also examined the subsequent correction to the electronic data files. We evaluated program conventions to determine whether they had been consistently and correctly used.

Standard Operating Procedures Because no documents could be identified as formal Standard Operating Procedures, the Audit Team reviewed and followed other less formal procedural documents and conventions. Ancillary documentation was largely limited to the Six Cities Study.

Conformance with Audit Standards The Audit Team worked in conjunction with the rest of the Reanalysis Team to identify variables for audit in the validation and sensitivity phases. We identified two random samples from the electronic data files from the investigators: 250 individuals from each study for auditing questionnaires and study population variables, and an additional 250 individuals from each study for auditing death certificates and vital status. We compared the most original form of data (ie, questionnaires, death certificates) to a printout of the study population variables for each random sample in each study to assure the accuracy of the information in the computer file and that source records supported the coding and entry of each variable. We compared the results of this checking procedure to the published information for each study.

Original Cohort Identification and City Selection Criteria The Audit Team's original intention was to inspect any records that explained criteria for including and excluding cohort members. Methods for selecting subjects in the Six Cities Study had been described in detail by Ferris and colleagues (1979). Table 1 of that publication (page 768) presented the methods for selecting subjects in each city. Subjects had been selected at random on the basis of household voting lists, private census lists, partial blocks from street lists, or alphabetized names lists.

The Audit Team could not evaluate subject selection criteria for the ACS Study because all selection of study subjects had been made by ACS volunteers who had been instructed to find subjects whom they could follow over the next 7 years. Therefore, volunteers picked relatives, neighbors, and friends whom they believed would fit the criterion of long-term follow-up. No records of selection criteria from volunteers could be found for auditing. Although we could not confirm this by our audit, it is likely that these volunteers selected individuals in similar socioeconomic groups as themselves.

Data Audit The Audit Team considered a statistically based audit to be the best approach. We adopted specific procedures from methods proposed originally by Siconolfi (1986). However, that publication did not provide sufficient details as to the statistical theory behind the proposed sampling approach. Rather than relying on published tables such as those provided by Siconolfi or Schilling and Sommers (1988), the Reanalysis Team performed sample-size calculations to be used by the Audit Team.

The purpose of these calculations was to determine the optimal size of the random sample that would ensure that the true error rate in the electronic data would fall within a certain acceptable limit. We audited data for all variables used in the original publications in both the verification and the sensitivity analysis stages. The goal of the sampling was to detect errors in each variable that would meaningfully impact the interpretation of the results of the regression analyses. On the basis of discussions with the Original Investigators, we learned that the data collection and quality control measures used in each study had not changed over time in any significant way. Therefore, the Audit Team concluded that stratifying the example by city or other variable was necessary.

The Reanalysis Team investigated several aspects of the sample size issue. First, we evaluated the probability of finding errors in the data when, in fact, errors do exist. For sample sizes that range from 10 to 250, Figure A.1 shows the probability (statistical power) of detecting at least one error in the sample as a function of the error rate in the study population. For a sample size of 250, the probability of detecting an error rate of 5% is close to 100%.

Second, the Reanalysis Team calculated the power of a sample of a given size to distinguish between an error rate of 5% or less and an error rate of greater than 5%. For sample sizes that range from 50 to 1,000, Figure A.2 shows the statistical power (on the ordinate) according to the true error rate in the sample (abscissa). This figure shows that the statistical power increases as the sample size increases, although sample sizes over 250 offer very little gain. We concluded from this that a sample size of about 250 should be adequate to distinguish between an error rate of 5% or less and an error rate of 10% or more. As indicated in Figure A.3, a sample size of 250 would also be able to distinguish

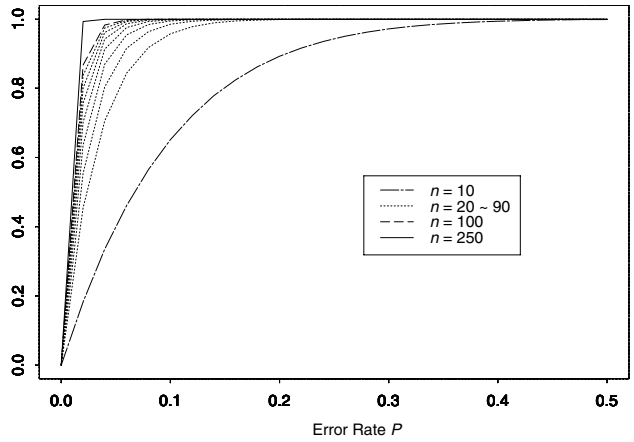

Figure A.1. Probability (statistical power) of detecting at least one error in a sample of size *n* as a function of the error rate in the study population.

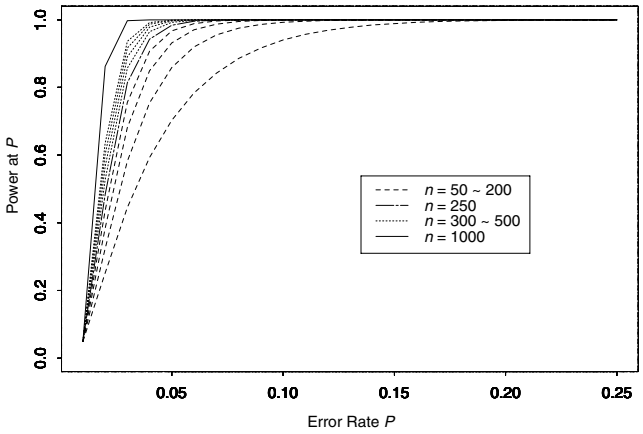

Figure A.3. Statistical power of different sample sizes to reject the null hypothesis of the error rate in the data being less than 1%.

between an error rate of 1% or less and error rates greater than 5% with high probability.

Third, the Reanalysis Team took a more classical approach to calculating sample size (Cochran 1977). We assumed five different levels of "statistical precision", defined as one half the CI (from 2% to 6%) at a level of statistical significance of 5%, and calculated what sample size would be necessary to achieve each of these levels of statistical precision. We considered error rates between 0% and 25%. We both included and excluded a term to correct for sampling from finite populations. Figure A.4 shows the results of these calculations. Sample sizes selected on the basis of a finite population were always larger, although very little difference is apparent for levels of statistical precision of 4% and higher.

Fourth, in addition to the above calculations, we also investigated the exact 95% CIs for sample sizes ranging from 200 to 500. The Reanalysis Team found that the CIs were very close for these sample sizes. For example, the 95% CIs for an error rate of 5% for sample sizes of 250 and 500 are 0.03–0.09% and 0.04–0.08%, respectively; for an error rate of 1%, the CIs are 0.00–0.04% and 0.00–0.03%, respectively. An important consideration in evaluating the significance of errors in the original variables is the impact of such errors on estimates of risk.

Results by Wang and colleagues (1994) for cohort mortality studies involving computerized record linkage suggest that the bias in risk estimates due to misascertainment of vital status may lead to biases in risk estimates proportional to the vital status error rate. Based on these results, the Reanalysis Team adopted a rule of thumb that error rates of less than 5% may not be of great epidemiologic importance.

In conclusion, these calculations show that a sample size of 250 was more than adequate for the purposes of this

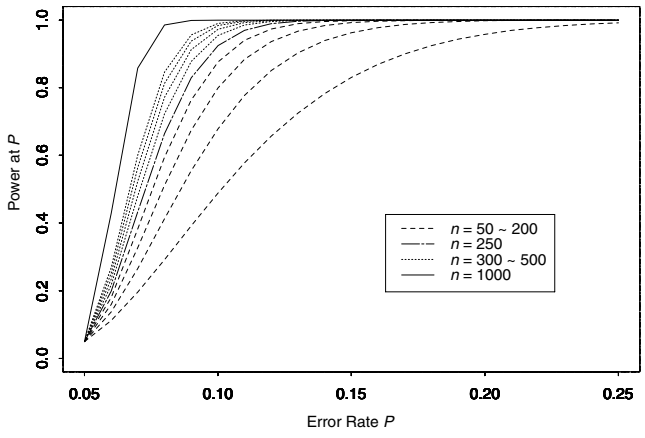

Figure A.2. Statistical power of different sample sizes to reject the null hypothesis of errors less than 5% in the data.

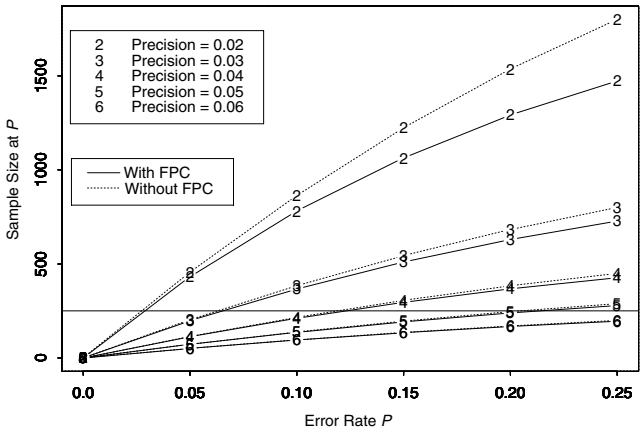

Figure A.4. Sample sizes needed to achieve different levels of statistical precision (0.02–0.06) depending on error rate *P*. Statistical significance was assumed to be 5%. Sample sizes were calculated both with and without a correction for sampling from a finite population (FPC).

data audit. For each of the two cohort studies, the Reanalysis Team randomly selected, without replacement, two samples of 250 each. (However, subjects selected in one sample could also be selected in the other.) We used one independent sample to audit death certificates, so that we sampled only those persons declared to be deceased in the original investigations. We selected the second independent sample from the entire cohort used in the published studies. Thus, it was possible for a subject to be included in both samples audited.

APPENDIX B. Analytic Methods of the Harvard Six Cities Study and the American Cancer Society Study

In these original investigations survival analysis with the Cox proportional-hazards model was used to estimate relative risks of mortality associated with air pollution. In this multivariate model, the ratio of the hazard function of the unexposed population to the exposed population provides an estimate of the relative risk. Formally, this multivariate 56

model is expressed as:

$$\lambda_A(t) = \lambda_B(t) \times \exp\left(\sum_{i=1}^{k} \beta_i X_i + \gamma E\right) \quad (1)$$

where $\lambda_A(t)$ and $\lambda_B(t)$ represent the hazard rates, as functions of time, in the exposed and baseline populations, respectively; X_i represents a series of potential confounding variables; and E represents the exposure to air pollutants. The coefficients β_i and γ, are estimated from the pseudolikelihood function and represent the natural logarithms of covariates. For example, $\exp(\gamma E)$ for E evaluated over the entire range (for sulfate in the ACS Study this was 19.9 $\mu g/m^3$) represents the mortality risk for an increase in exposure across the range of E. Both the ACS and Six Cities studies used the Cox proportional-hazards model to estimate risk. The variables used in each study are outlined separately below.

AMERICAN CANCER SOCIETY STUDY

The manuscript by the original investigators (Pope et al 1995) describes the methods used in detail. The Cox proportional-hazards model was applied using the survival time from the date of enrollment. The survival times of those who had not died were censored at the end of the study's follow-up period. The Cox proportional-hazards models were stratified by 5-year age groupings, sex, and race, which permitted a baseline hazard, $\lambda_B(t)$, to be estimated within each such stratum. A stratified Cox proportional-hazards model can be expressed by extending the formula presented in Equation 1 to read

$$\frac{\lambda_{A_s}(t)}{\lambda_{B_s}(t)} \quad (2)$$

where s represents the strata defined by one or more categorical variables. This model includes separate, but not necessarily proportional, hazards for each stratum.

Separate models were fit for two air pollution variables: the mean concentration of sulfate particle pollution for 1980 in the participant's place of residence, and the median fine particle concentration for 1979–1983 calculated from the dichotomous sampler network by Lipfert and colleagues (1988). Scaling was applied so that the parameter estimates would yield a relative risk for the most-polluted area relative to the least-polluted area. For sulfate particles, this factor was 19.9 $\mu g/m^3$ and for fine particles it was 24.5 $\mu g/m^3$.

A comprehensive listing of potential confounding variables was entered into the multivariate model. The following variables were included to adjust for smoking behavior: an indicator variable for current-smokers; an indicator variable for pipe smokers, cigar smokers, or both; number of years smoked for current-smokers; cigarettes smoked daily for current-smokers; years smoked for former-smokers; cigarettes smoked daily for former-smokers; and number of hours per day passively exposed to smoke. Other risk factors that were controlled for in the analyses included BMI, drinks per day of alcohol, a dichotomous variable indicating whether high school education had been attained or not, and variables representing occupational exposure to any of several substances (asbestos, chemicals or acids or solvents, coal or stone dusts, coal tar or pitch or asphalt, diesel engine exhaust, or formaldehyde).

Cox proportional-hazards models were used to derive risk estimates for lung cancer (ICD-9 code 162), cardiopulmonary diseases (ICD-9 codes 401–440), and all causes of death. Risks were also calculated for current-smokers relative to never-smokers under the assumption that current-smokers smoked 20 cigarettes per day for a period of 25 years.

HARVARD SIX CITIES STUDY

Similar methods were used in the Six Cities analyses (Dockery et al 1993). Cox proportional-hazards modeling was used with stratification by 5-year age groupings and sex. The series of risk factors included in these models were indicator variables for current-smokers and former-smokers; number of pack-years of smoking (for current-smokers and former-smokers separately); an indicator variable for having attained high school education or not; a

continuous-measure BMI; and binary variables denoting exposure to dusts, gases, or fumes.

Mean concentrations of fine particles and the city of residence were used as separate indicators of air pollution exposure. That means the relative risk of mortality due to air pollution exposure was evaluated in two ways. First, indicator variables were created for each city of residence by estimating the relative risk of mortality for each city using Portage, the city with the lowest concentration of fine particle air pollution, as the reference category. These relative risks by city of residence were presented separately for males and females.

Second, the Cox proportional-hazards model was used to estimate the relative risk of mortality using a continuous measure of the concentration of fine particles that included all fine particle data regardless of city or year. The logarithm estimates of relative risk were multiplied by 18.6 $\mu g/m^3$ in order to provide an estimate of the relative risk for residents of the most-polluted city (Steubenville OH) relative to residents of the least-polluted city (Portage WI).

Cox proportional-hazards models were fitted for four cause-of-death categories: all causes, lung cancer (ICD-9 code 162), cardiopulmonary disease (ICD-9 codes 400–440, 485–496), and all other causes. In a separate group-specific analysis, the Original Investigators calculated mortality risks for current-smokers, assuming these individuals had accrued 25 pack-years of smoking compared with never-smokers. Similarly, risks for former-smokers were calculated assuming 20 pack-years of smoking.

APPENDIX C. Questionnaires and Codebook Used in the Harvard Six Cities Study

Form 78 (1/77)

1

HARVARD QUESTIONNAIRE on RESPIRATORY SYMPTOMS (Adults)

I.D. (1-4) ☐☐☐☐☐ (1 2 3 4 5)

Card No.: ☐ (6)

NAME _____ (Last / First)

Soc. Sec. No.: _____ ☐☐☐☐☐☐ (7 8 9 10 11 12)

Date of Birth: _____ (7-12) Mo. Day Yr.

Tel. No. _____

Place of Birth: _____ (13-14) ☐☐ (13 14)

Address: _____

Sex (15) Male _____ 1 Female _____ 2 ☐ (15)

_____ City Zip

(To be coded later): S.C. _____ (16) ☐ (16)

Occupation Code: _____ (17-18-19) ☐☐☐ (17 18 19)

(20) **MARITAL STATUS:**

Single _____ 1
Married _____ 2
Widowed _____ 3
Sep/Div. _____ 4

(21) **RACE:**

Caucasian _____ 1
Negro _____ 2
Oriental _____ 3
Am. Indian/ Mexican _____ 4
Other _____ 5

☐ (20)

☐ (21)

SPOUSE:

Name: _____

Study No. _____ (22-25) ☐☐☐☐ (22 23 24 25)

Occupation: _____

OCCUPATIONAL HISTORY (from time of leaving school)

Residences with dates	Industry	Actual Job	Materials Handled	Years of Exposure Dust	Years of Exposure Gas/Fumes
Presently:					
			TOTALS	(26-27)	(28-29)

☐☐ (26 27)

☐☐ (28 29)

(Be sure to get current occupation ("retired" etc.))

(30) Number of job changes in last 10 years _____ ☐ (30)

(31) Number of residence changes (change of town) in last 10 years _____ ☐ (31)

(32-33) Number of years resident in this town _____ ☐☐ (32 33)

(34-35) Years in same part of town _____ ☐☐ (34 35)

(36)
Highest Grade of Schooling Completed:

Grade school not completed	____ 0	Trade school or only attended college	____ 3
Grade school completed	____ 1	2-yr. college or nursing graduate	____ 4
High school completed	____ 2	4-yr. college graduate	____ 5
		Post-graduate	____ 6

☐ 36

CURRENT HEALTH *("YES" must satisfy criteria of ". . . as much as 4-6 times a day for 4 days of the week")*

Cough

(37) A. Do you usually have a cough? No ____ 0 Yes ____ 1 ☐ 37

(38) B. Do you cough at all on getting up, or first thing in the morning?

In winter: No ____ Yes ____ 1
In summer: No ____ Yes ____ 2
(Add "Yes" scores) ☐ 38

(39) C. Do you go on coughing during the day or at night?

In winter: No ____ Yes ____ 1
In summer: No ____ Yes ____ 2
(Add "Yes" scores) ☐ 39

If YES to A, B, or C ask:

D. Do you cough like this on most days for as much as 3 months at a time?

(40) No ____ 0 Yes ____ 1

E. How long have you had this cough?
Less than 3 years ____ 0
3 years or more ____ 2
N/A ____ 8
(Add "Yes" scores) ☐ 40

Phlegm

(41) A. Do you usually bring up phlegm from your chest (not from the back of your nose)?

No ____ 0 Yes ____ 1 ☐ 41

(42) B. Do you bring up phlegm at all on getting up, or first thing in the morning?

In winter: No ____ 0 Yes ____ 1
In summer: No ____ 0 Yes ____ 2
(Add "Yes" scores) ☐ 42

(43) C. Do you go on bringing up phlegm during the day?

In winter: No ____ 0 Yes ____ 1
In summer: No ____ 0 Yes ____ 2
(Add "Yes" scores) ☐ 43

If YES to A, B, or C ask:

D. Do you bring up phlegm like this on most days for as much as three months at a time?

(44) No ____ 0 Yes ____ 1

E. How long have you had this phlegm?
Less than 3 years ____ 0
3 years or more ____ 2
N/A ____ 8
(Add "Yes" scores) ☐ 44

(45) F. Do you get bouts of (increased) cough and phlegm lasting for 3 weeks each winter?

No ____ 0 Yes, last 3 winters only ____ 1 Yes, more than 3 winters ____ 2 ☐ 45

Wheezing

(46) A. Does your chest ever sound wheezy or whistling?

If NO, ask: Not even when you have a cold? No ____0 Yes ____1

If YES, ask: Only with colds? ____ Occasionally apart from colds? ____2

Most days or nights? ____3

(Record highest score)

If YES to most days or nights, ask:

Has this been present for the past 3 years or more? No ____ Yes ____4

46

(47) B. Have you ever had an attack of wheezing that has made you feel short of breath?

No ____0 Yes ____1

If YES, ask:

Have you had 2 or more such episodes? No ____0 Yes ____1

Have you ever required medicine or treatment for the(se) attack(s)? No ____0 Yes ____2

(Add "Yes" scores)

47

(48) **Breathlessness**

If disabled from walking by any condition other than heart or lung disease, describe and do not ask the following questions (A-F): ____8

Are you troubled by shortness of breath?

A. If NO, ask: Not even when hurrying on the level or walking up a slight hill?

No ____0 Yes ____1

B. If YES, ask: Do you have to walk slower than people of your age on the level because of breathlessness?

No ____ Yes ____2

C. If YES, ask: Do you ever have to stop for breath when walking at your own pace on the level?

No ____ Yes ____3

D. If YES, ask: Do you have to stop for breath after walking about 100 yards (or after a few minutes) on the level?

No ____ Yes ____4

E. If YES, ask: Are you too breathless to leave the house, or breathless on dressing or undressing?

No ____ Yes ____5

(Record highest score)

48

(49) **If Yes to C ask:**

Have these symptoms been present for at least the past 2 winters? No ____0 Yes ____1 N/A ____8

49

(50) **Colds**

If you get a cold, does it usually go to your chest? No ____0 Yes ____1

50

Effect of Weather

(51) A. Does the weather affect your chest or breathing? No ____ 0 Yes ____ 1 [51]

If YES, ask B, C, and D.

(52) B. Does foggy or damp weather affect it? No ____ Yes ____ 1

C. Does cold weather affect it? No ____ Yes ____ 2 *(Add "Yes" scores)* [52]

D. Does hot weather affect it? No ____ Yes ____ 4

N/A ____ 8

(If YES to B, C, or D, ask E, F, and G.)

(53) E. Does this weather make you short of breath? No ____ Yes ____ 1

F. Does this weather make you wheeze? No ____ Yes ____ 2 *(Add "Yes" scores)* [53]

G. Does this weather increase your cough or phlegm? No ____ Yes ____ 4

N/A ____ 8

(54) **PAST CHEST ILLNESSES**

During the past 3 years have you had any chest illness that has kept you off work, indoors at home or in bed? No ____ 0 Yes ____

If YES, ask details of each illness; if NO, ask: "Not even flu?"

Year	Lasted 1 week or more? No	Lasted 1 week or more? Yes	Had increased phlegm? No	Had increased phlegm? Yes	Diagnosis	Score

If YES in both columns, Score 1. Total Score (54) [54]

Has a doctor ever said you had:

		No	Yes
(55)	Bronchitis	0	1
	Emphysema	0	2
	Pneumonia	0	4

(Add "Yes" scores) [55]

		No	Yes
(56)	Sinus Trouble	0	1
	Pulmonary Tuberculosis	0	2
	Hay Fever	0	4

(Add "Yes" scores) [56]

(57) Bronchial Asthma No ____ 0 Yes, at present ____ 1 Yes, in past (but not now) ____ 2 [57]

(58) Other chest illness No ____ 0 Yes ____ 1 Specify _____

chest operations No ____ 0 Yes ____ 2 Specify _____

chest injury No ____ 0 Yes ____ 4 Specify _____

(Add "Yes" scores) [58]

5

PAST ILLNESS – GENERAL

(59) Has a doctor ever said you had diabetes or have you been told you had sugar in your urine (water) or too much sugar in your blood?
No ____ 0 Yes ____

If YES: Are you currently taking medication -
by injection ..... Yes ____ 1
or taking medication by mouth ..... Yes ____ 2
or controlled only by diet ..... Yes ____ 3
or none of these ..... Yes ____ 4
(Record score) → 59

(60) Has a doctor ever told you that you had heart trouble? No ____ 0 Yes ____ 1
If YES, ask: Have you had any *treatment* for it in the past 10 years? No ____ 0 Yes ____ 1
Has a doctor ever told you that your blood pressure was high? No ____ 0 Yes ____ 3
If YES, ask: Have you had any *treatment* for it in the past 10 years? No ____ 0 Yes ____ 3
(Add "Yes" scores) → 60

(61) Have you been bothered by indigestion, heartburn, or pains in the stomach in the past 10 years? No ____ 0 Yes ____ 1
If YES:
Has this been as much as twice a week for as long as a month? No ____ 0 Yes ____ 1
(Add "Yes" scores) → 61
Doctor's diagnosis, if given _____

(62) Have you ever vomited any blood? No ____ 0 Yes ____ 1
Have you ever been told you had stomach ulcers? No ____ 0 Yes ____ 2
(Add "Yes" scores) → 62

Have you ever had an operation on the abdomen or stomach for:

(63)

	No	Yes
Ulcers	____ 0	____ 1
Appendix	____ 0	____ 2
Gallbladder	____ 0	____ 4

(Add "Yes" scores) → 63

(64)

	No	Yes
Exploratory Laparotomy	____ 0	____ 1
Other Intestinal	____ 0	____ 2
_____		

(Add "Yes" scores) → 64

(65) Are you troubled by frequent headaches? No ____ 0 Yes ____ 1 → 65

(66) Have you ever been affected by gas or fumes at work or anywhere? No ____ 0 Yes ____
If YES, ask: Once or twice? ____ 1 More often? ____ 2 *(Record score)* → 66

(67) For any occurrence ask: Did you ever have to see a doctor? No ____ 0 Yes ____
If YES, ask: Once? ____ 1 More than once? ____ 2 N/A ____ 8 *(Record score)* → 67

(68) If doctor was seen, ask: Were you hospitalized for a day or more? No ____ 0 Yes ____
If YES, ask: Once? ____ 1 More than once? ____ 2 N/A ____ 8 *(Record score)* → 68

78 79 7 80 8

CARD 2

I.D. (1-5)

6

Card No.: 2 (6)

DRINKING

(7) A. Do you PRESENTLY use alcoholic beverages? No ____ 0 Yes ____ 1

B. If YES, Is this as often as 1 day per week? No ____ 0 Yes ____ 1

(Add "yes" scores) (7)

C. *If YES to B, ask for each of beer, wine and liquor:*

"How much do you drink on an average per week?"

(8) Beer (oz./wk.)	(9) Wine (oz./wk.)	(10) Liquor (oz./wk.)
None ____ 0	None ____ 0	None ____ 0
1-199 ____ 1	1-99 ____ 1	1-25 ____ 1
200+ ____ 2	100+ ____ 2	26+ ____ 2

(8) (9) (10)

[N.B.: 200 oz. = 25 8-oz. glasses 16 oz. = 1 pint
100 oz. = 25 4-oz. glasses 26 oz. = 1/5 gallon]

TOBACCO SMOKING

(11) Have you *ever* smoked? No ____ 0 Yes ____ 1

Do you *now* smoke? No ____ 0 Yes ____ 2

(As of 1 month ago)

(Record highest score) (11)

("No" means less than 20 packs of cigarettes or 12 oz. of tobacco in a lifetime.)

For current or ex-smokers, ask:

(12) Have you ever been able to stop smoking cigarettes for 6 months or longer? No ____ 1 Yes ____ (12)

If YES:

Did you practice total abstinence? No ____ Yes ____ 2

Did you switch to cigar/pipe? No ____ Yes ____ 3

Did you chew or take snuff? No ____ Yes ____ 4

N/A ____ 8

(Record score checked; if both 3 and 4 checked, record 5)

For current and ex-smokers, obtain the following:

Duration of Smoking

	Age Started*	Age finally stopped; if not, current age:	Years Abstinence	Total Yrs. Smoked*
Cigarettes	(13-14)			(15-16)
Pipe				(17-18)
Cigars				(19-20)

(13) (14) (15) (16) (17) (18) (19) (20)

Enter 00 for Never Smoked.

7

AMOUNT SMOKED

Packs of (20) Cigarettes/week

	(a) (21-22) Weekly Amt.	(b) Years Duration	(a) x (b)
Present Pattern			
Past Periods			
Total			(25-28)

Hand-rolled tobacco oz./wk.

	(c) (23-24) Weekly Amt.	(d) Years Duration	(c) x (d)
Total			(29-32)

Pipe Tobacco oz./wk.

	(e) (33-34) Weekly Amt.	(f) Years Duration	(e) x (f)
Present Pattern			
Past Periods			
Total			(37-40)

Cigars/wk.

	(g) (35-36) Weekly Amt.	(h) Years Duration	(g) x (h)
Total			(41-44)

(45) Do you smoke filter-tip cigarettes currently? *(Currently means for past month at least.)*

Never ____ 1 Less than ½ time ____ 2 ½ time + ____ 3 Always ____ 4 N/A ____ 8

How long have you been using filter-tips? ____ years

	Slightly	Moderately	Deeply	Not at all	Never smoked
(46) Do/did you inhale the cigarette smoke?	____ 1	____ 2	____ 3	____ 4	____ 5
(47) Do/did you inhale the pipe/cigar smoke?	____ 1	____ 2	____ 3	____ 4	____ 5

(48) Have you ever chewed tobacco regularly? No ____ 0 Yes ____ 1

Have you ever used snuff regularly? No ____ 0 Yes ____ 2

(Add "Yes" scores)

21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48

8

ALLERGIES

For each of the following, if response is "Yes," ask whether confirmed by doctor

		None	Confirmed by Doctor: No	Confirmed by Doctor: Yes
Have you ever had an allergic reaction to				
(49)	food or medicine (ingested)?	0	1	2
(50)	pollen, dust (inhaled)?	0	1	2
(51)	detergents, metals (skin contact)?	0	1	2

□ □ □ 49 50 51

FAMILY HISTORY

Were either of your natural parents ever told by a doctor that they had a chronic lung condition such as:

		Father			Mother		
		No (0)	Yes (1)	Don't know (2)	No (0)	Yes (1)	Don't Know (2)
(52)	chronic bronchitis?						
(53)	emphysema?						
(54)	asthma?						
(55-57)	other lung disease? (specify)						
					(58)		
					(59)		
					(60)		
					(61-63)		

□ □ □ 52 53 54

□ □ □ 55 56 57

□ □ □ 58 59 60

□ □ □ 61 62 63

(64) PULMONARY FUNCTION

Test done _____ 1

Test not done because:

Subject could not perform test _____ 2

Refused _____ 3

Other, specify _____ 4

□ 64

(65) SELECTION: Normal _____ 0

Volunteer No _____ Yes _____ 1

Interpreter No _____ Yes _____ 2

At Home No _____ Yes _____ 4

□ 65

(66-67) INTERVIEWER _____

□ □ 66 67

(68-73) DATE OF INTERVIEW _____
Month Day Year

□ □ □ □ □ □ 68 69 70 71 72 73

□ □ □ □ □ 74 75 76 77 78

7 8 79 80

Form 82 (8/81)

1

HARVARD QUESTIONNAIRE on RESPIRATORY SYMPTOMS (Adults)
(FOLLOW-UP)

I.D. (1-4) ☐☐☐☐☐ (1 2 3 4 5)

Card No.: ☐ (6)

NAME _____ (Last) (First)

Soc. Sec. No.: _____ ☐☐☐☐☐☐ (7 8 9 10 11 12)

Date of Birth: (7-12) _____ Mo. Day Yr.

Tel. No. _____

Sex (13) Male ____ 1 Female ____ 2

Current Address: _____ ☐☐☐ (13 14 15)

SPOUSE:

Name: _____

_____ City Zip (16-20) ☐☐☐☐☐ (16 17 18 19 20)

Study No. _____

(21) **MARITAL STATUS:** ☐☐ (21 22)

Single ____ 1 Widowed ____ 3

Married ____ 2 Sep/Div. ____ 4

(22) **RACE:**

Caucasian ____ 1 Oriental ____ 3

Black ____ 2 Am. Indian/ Mexican ____ 4

Other ____ 5

OCCUPATIONAL HISTORY (During the past three years)

Spouse Job Title: _____ Spouse Industry: _____

Residences with dates	Industry	Actual Job	Materials Exposed To	Years of Exposure: Dust	Years of Exposure: Gas/Fumes
Presently:	(23-25)	(26-28)			
			TOTALS	(29-30)	(31-32)

☐☐☐ (23 24 25) ☐☐☐ (26 27 28) ☐☐ (29 30) ☐☐ (31 32)

Be sure to obtain both present industry and present job (actual work activity).

(33) Number of job changes in last 3 years: ____ ☐ (33)

(34) Have you changed your address in the last three years? No ____ Yes ____ 1 ☐ (34)

If Yes, do you live in the same part of town as you did three years ago? No ____ Yes ____ 2
(Add "Yes" scores)

(35-36) How long have you lived at your current address? ____ yrs ☐☐ (35 36)

(37) What fuel do you use most to heat your home? (One only)

Oil ____ 1 Natural gas ____ 2 Bottled gas (Propane) ____ 3 Electricity ____ 4 Wood ____ 5 Coal ____ 6

Other, (Specify) _____ 7 ☐ (37)

(38-39) What other fuels are used to heat your home? (Check all appropriate.)

Oil ____ 1 Natural gas ____ 2 Bottled gas (Propane) ____ 4 Electricity ____ 8 Wood ____ 16 Coal ____ 32

Other, (Specify) _____ 0 (Add Scores) ☐☐ (38 39)

2

(40) How is heat mainly distributed to the rooms in your home? (One only)

Hot Air ___1 Hot Water ___2 Steam ___3 Fireplace ___4 Stove ___5 Space Heater ___6

Other, (Specify) _____ 7 ☐ 40

(41-42) What other heat distribution methods are used in your house? (Check all appropriate)

Hot Air ___1 Hot Water ___2 Steam ___4 Fireplace ___8 Stove ___16 Space Heater ___32

Other, (Specify) _____ 0 (Add Scores) ☐☐ 41 42

(43-44) Have you made any attempts to seal up your house for energy conservation in the last 6 years?

No ___0 Yes ___

No, house was already tight ___1 Storm windows ___2

Don't know or not applicable ___99 Caulking ___4

Insulation ___8

Flue Damper ___16

Other _____ 32 (Add Scores) ☐☐ 43 44

(45) What fuel is used for cooking? Gas ___ Coal ___ Wood ___ Electricity ___ Other _____ ☐ 45

(46) Do you have an exhaust fan for your cooking stove? No ___ Yes ___ Don't know ___ ☐ 46

(47) If yes, do you use it: Never ___ Seldom ___ Regularly ___ ☐ 47

(48) How is the exhaust fan vented? into the room ___; vented to outside ___; Don't know ___ ☐ 48

(49) Is your home air-conditioned? No ___; Yes, partially ___; Yes, completely ___ ☐ 49

The following questions apply to your HEALTH in the last three years.

Please answer Yes or No if possible.

CURRENT HEALTH *("YES" must satisfy criteria of ". . . as much as 4-6 times a day for 4 days of the week")*

Cough.

(50) A. Do you usually have a cough? No _____ 0 Yes _____ 1 ☐ 50

(51) B. Do you cough at all on getting up, or first thing in the morning?

In winter: No _____ Yes _____ 1

In summer: No _____ Yes _____ 2

(Add "Yes" scores) ☐ 51

(52) C. Do you go on coughing during the day or at night?

In winter: No _____ Yes _____ 1

In summer: No _____ Yes _____ 2

(Add "Yes" scores) ☐ 52

If YES to A, B, or C ask:

D. Do you cough like this on most days for as much as 3 months at a time?

(53) No _____ 0 Yes _____ 1

E. How long have you had this cough? Less than 3 years _____ 0

3 years or more _____ 2

N/A _____ 8

(Add "Yes" scores) ☐ 53

Phlegm

(54) A. Do you usually bring up phlegm from your chest (not from the back of your nose)?

No ____ 0 Yes ____ 1 □ 54

(55) B. Do you bring up phlegm at all on getting up, or first thing in the morning?

In winter: No ____ 0 Yes ____ 1
In summer: No ____ 0 Yes ____ 2
(Add "Yes" scores) □ 55

(56) C. Do you go on bringing up phlegm during the day?

In winter: No ____ 0 Yes ____ 1
In summer: No ____ 0 Yes ____ 2
(Add "Yes" scores) □ 56

If YES to A, B, or C ask:

(57) D. Do you bring up phlegm like this on most days for as much as three months at a time?

No ____ 0 Yes ____ 1

E. How long have you had this phlegm?

Less than 3 years ____ 0
3 years or more ____ 2
N/A ____ 8
(Add "Yes" scores) □ 57

(58) F. Do you get bouts of (increased) cough and phlegm lasting for 3 weeks each winter?

No ____ 0 Yes, last 3 winters only ____ 1 Yes, more than 3 winters ____ 2 □ 58

Wheezing

(59) A. During the past three years, has your chest ever sounded wheezy or whistling?

If NO, ask: Not even when you have a cold? No ____ 0 Yes ____ 1

If YES, ask: Only with colds? ____ Occasionally apart from colds? ____ 2
Most days or nights? ____ 3

If YES to most days or nights, ask:
Has this been present for the past 3 years or more? No ____ Yes ____ 4
(Record highest score) □ 59

(60) B. During the past three years, have you had an attack of wheezing that has made you feel short of breath?

No ____ 0 Yes ____ 1

If YES, ask:
Have you had 2 or more such episodes? No ____ 0 Yes ____ 1
Have you ever required medicine or treatment for the(se) attack(s)? No ____ 0 Yes ____ 2
(Add "Yes" scores) □ 60

(61) **Breathlessness**

If disabled from walking by any condition other than heart or lung disease, describe and do not ask the following questions (A-F): ____ 8

Are you troubled by shortness of breath?

A. If NO, ask: Not even when hurrying on the level or walking up a slight hill?

No ____ 0 Yes ____ 1

B. If YES, ask: Do you have to walk slower than people of your age on the level because of breathlessness?

No ____ Yes ____ 2

C. If YES, ask: Do you ever have to stop for breath when walking at your own pace on the level?

No ____ Yes ____ 3

D. If YES, ask: Do you have to stop for breath after walking about 100 yards (or after a few minutes) on the level?

No ____ Yes ____ 4

E. If YES, ask: Are you too breathless to leave the house, or breathless on dressing or undressing?

No ____ Yes ____ 5

(Record highest score) □ 61

4

(62) **If Yes to C ask:**

Have these symptoms been present for at least the past 2 winters? No ____ 0 Yes ____ 1 N/A ____ 8 [62]

(63) **Colds**

If you get a cold, does it usually go to your chest? No ____ 0 Yes ____ 1 [63]

Effect of Weather

(64) A. Does the weather affect your chest or breathing? No ____ 0 Yes ____ 1 [64]

If YES, ask B, C, and D.

B. Does foggy or damp weather affect it? No ____ Yes ____ 1

(65) C. Does cold weather affect it? No ____ Yes ____ 2 *(Add "Yes" scores)* [65]

D. Does hot weather affect it? No ____ Yes ____ 4

N/A ____ 8

(If YES to B, C, or D, ask E, F, and G.)

E. Does this weather make you short of breath? No ____ Yes ____ 1

(66) F. Does this weather make you wheeze? No ____ Yes ____ 2 *(Add "Yes" scores)* [66]

G. Does this weather increase your cough or phlegm? No ____ Yes ____ 4

N/A ____ 8

(67) **PAST CHEST ILLNESSES**

During the past 3 years have you had any chest illness that has kept you off work, indoors at home or in bed? No ____ 0 Yes ____

If YES, ask details of each illness; if NO, ask: "Not even flu?"

Year	Lasted 1 week or more? No	Lasted 1 week or more? Yes	Had increased phlegm? No	Had increased phlegm? Yes	Diagnosis	Score
					.....	
					.....	
					.....	

If YES in both columns, Score 1. Total Score

(67) [67]

During the past three years, has a Doctor said you have:

		No	Yes
(68)	Bronchitis	0	1
	Emphysema	0	2
	Pneumonia	0	4

(Add "Yes" scores) [68]

		No	Yes
(69)	Sinus Trouble	0	1
	Pulmonary Tuberculosis	0	2
	Hay Fever	0	4

(Add "Yes" scores) [69]

5

(70) Bronchial Asthma No ____0 Yes, at present ____1 Yes, in past (but not now) ____2 [70] [71]

(71) Other chest illness No ____0 Yes ____1 Specify ____

chest operations No ____0 Yes ____2 Specify ____

chest injury No ____0 Yes ____4 Specify ____

(Add "Yes" scores)

PAST ILLNESS – GENERAL

(72) Has a doctor ever told you that you had heart trouble? No ____0 Yes ____1

If YES, ask: Have you had any *treatment* for it in the past 3 years? No ____0 Yes ____1

Has a doctor ever told you that your blood pressure was high? No ____0 Yes ____3

If YES, ask: Have you had any *treatment* for it in the past 3 years? No ____0 Yes ____3 [72]

(Add "Yes" scores)

(73) Are you troubled by frequent headaches? No ____0 Yes ____1 [73]

(74) During the past three years, have you been affected by gas or fumes at work or anywhere? No ____0 Yes ____

If YES, ask: Once or twice? ____1 More often? ____2 *(Record score)* [74]

(75) For any occurrence ask: Did you ever have to see a doctor? No ____0 Yes ____

If YES, ask: Once? ____1 More than once? ____2 N/A ____8 *(Record score)* [75]

(76) If doctor was seen, ask: Were you hospitalized for a day or more? No ____0 Yes ____

If YES, ask: Once? ____1 More than once? ____2 N/A ____8 *(Record score)* [76] [77] [78]

[5] [4] [79] [80]

CARD 2

I.D. (1-5)

6 | 1 2 3 4 5

Card No.: 2 | 6

TOBACCO SMOKING

(7) Have you *ever* smoked? No ____ 0 Yes ____ 1

Do you *now* smoke? No ____ 0 Yes ____ 2 *(Record highest score)*

(As of 1 month ago)

("No" means less than 20 packs of cigarettes or 12 oz. of tobacco in a lifetime.)

For current or ex-smokers, ask:

(8) Have you ever been able to stop smoking cigarettes for 6 months or longer? No ____ 1 Yes ____

7 | 8

If YES:

Did you practice total abstinence? No ____ Yes ____ 2

Did you switch to cigar/pipe? No ____ Yes ____ 3

Did you chew or take snuff? No ____ Yes ____ 4 N/A____ 8

(Record score checked; if both 3 and 4 checked, record 5)

For current and ex-smokers, obtain the following:

Duration of Smoking

	Age Started*	Age finally stopped; if not, current age:	Years Abstinence	Total Yrs. Smoked*
Cigarettes	(9-10)			(11-12)
Pipe				(13-14)
Cigars				(15-16)

9 | 10 | 11 | 12

13 | 14 | 15 | 16

Enter 00 for Never Smoked.

AMOUNT SMOKED IN LAST THREE YEARS

Packs of (20) Cigarettes/week

(17-18)

	(a) Weekly Amt.	(b) Years Duration	(a) x (b)
Present Pattern			
Past Three Years			
Total			(21-24)

17 | 18

19 | 20

21 | 22 | 23 | 24

Hand-rolled tobacco oz./wk.

(19-20)

	(c) Weekly Amt.	(d) Years Duration	(c) x (d)
Total			(25-28)

25 | 26 | 27 | 28

Pipe Tobacco oz./wk.

(29-30)

	(e) Weekly Amt.	(f) Years Duration	(e) x (f)
Present Pattern			
Past Three Years			
Total			(33-36)

29 | 30

31 | 32

33 | 34 | 35 | 36

Cigars/wk.

(31-32)

	(g) Weekly Amt.	(h) Years Duration	(g) x (h)
Total			(37-40)

37 | 38 | 39 | 40

(41) Do you smoke filter-tip cigarettes currently? *(Currently means for past month at least.)*

Never ____1 Less than ½ time ____2 ½ time + ____3 Always ____4 N/A ____8

How long have you been using filter-tips? ____ years

☐ 41

		Slightly	Moderately	Deeply	Not at all	Never smoked
(42)	Do/did you inhale the cigarette smoke?	____ 1	____ 2	____ 3	____ 4	____ 5
(43)	Do/did you inhale the pipe/cigar smoke?	____ 1	____ 2	____ 3	____ 4	____ 5

☐ 42

☐ 43

During the last three years,

(44) Have you ever chewed tobacco regularly? No ____0 Yes ____1

Have you ever used snuff regularly? No ____0 Yes ____2

(Add "Yes" scores)

☐ 44

(45) How many persons 14 years and older live in your home? (Include yourself) ____

☐ 45

(46) How many of these are smokers?

☐ 46

(47) Are you normally exposed to other smokers away from home; for example, at work? No ____

Yes, occasionally ____1 Regularly ____2

☐ 47

If YES, is your exposure to smoke at these times: Light ____0 Moderate ____4 Heavy ____6

(Add Scores)

ALLERGIES (LIFETIME)

For each of the following, if response is "Yes," ask whether confirmed by doctor

During your entire lifetime have you ever had an allergic reaction to	None	*Confirmed by Doctor* No	Yes
(48) food or medicine or anything else you have ingested?	____ 0	____ 1	____ 2
(49) pollen, dust or anything else you inhaled?	____ 0	____ 1	____ 2
(50) detergents, metals or anything else you touched?	____ 0	____ 1	____ 2

☐ 48 ☐ 49 ☐ 50

FAMILY HISTORY

Were either of your natural parents ever told by a doctor that they had a chronic lung condition such as:

		Father			Mother		
		No (0)	Yes (1)	Don't know (2)	No (0)	Yes (1)	Don't Know (2)
(51)	chronic bronchitis?	____	____	____	(55) ____	____	____
(52)	emphysema?	____	____	____	(56) ____	____	____
(53)	asthma?	____	____	____	(57) ____	____	____
(54)	other lung disease? (specify)	____	____	____	(58) ____	____	____
		_____			_____		

☐ 51 ☐ 52 ☐ 53 ☐ 54

☐ 55 ☐ 56 ☐ 57 ☐ 58

(59) PULMONARY FUNCTION

Test done ____ 1

Test not done because:

Subject could not perform test ____ 2

Refused ____ 3

Other, specify _____ 4

Test done, questionable reading _____ 5
(specify)

☐ 59

8

(60) SELECTION: Normal ____ 0

Interpreter No ____ Yes ____ 2

At Home No ____ Yes ____ 4

(61-62) INTERVIEWER _____

(63-68) DATE OF INTERVIEW _____
Month Day Year

(69-70) Air Pollution Zones

60 61 62

63 64 65 66 67 68

69 70

5 79 4 80

I.D.

SPOUSE I.D.

We are attempting to identify children in the school survey who live in households with adults in this survey. If there are any children living in your home between the ages of ____ and ____, please give the following information:

NAME	SCHOOL	GRADE

NONE _____

This may be the last time we see you, but we would like to be able to keep you up to date on the results of the study. Is there someone who would know your new address if you move?

NAME _____ RELATIONSHIP _____

ADDRESS _____

Is there another person who would know your new address?

NAME _____ RELATIONSHIP _____

ADDRESS _____

Six Cities Reanalysis
Codebook for mort6c

Variable name	*Meaning and response codes*
1) ID	5-digit ID; last digit is city code
2) FVC1	Forced vital capacity in liters
3) FEV11	Forced expiratory volume in 1 second in liters
4) CURRCIG1	Current cigarette smoker(0 = No, 1 = Yes)
5) DUST_TO1	Total years of occupational dust exposure
6) EDUC11	Highest grade of schooling completed 0 = Grade school not completed 1 = Grade school completed 2 = High school completed 3 = Trade school or only attended college 4 = 2-yr college or nursing graduate 5 = 4-yr college graduate 6 = Postgraduate
7) FUM_TOT1	Total years of occupational gas or fumes exposure
8) SEX1	Sex (0 = Male, 1 = Female)
9) CIG_DAY1	Cigarettes per day
10) HEIGHT1	Height in meters
11) PKYRS1	Pack years of smoking
12) WEIGHT1	Weight in pounds
13) AGE_1	Age in years
14) CITY	City 1 = Watertown 2 = Kingston/Harriman 4 = St. Louis 5 = Steubenville 6 = Portage 9 = Topeka
15) INITDATE	Date of initial visit
16) LASTDATE	Last date on study for mortality analysis. If patient was dead, this date is the date of death; otherwise, this date is the last date we were able to confirm the patient was alive.
17) DEAD	Dead 0 = No (alive at LASTDATE), 1 = Yes
18) TIMEON	Time on study in years
19) DIAB	Diabetes (0 = No, 1 = Yes)
20) HI_BP	High blood pressure (0 = No, 1 = Yes)
21) CODE	Cause of death alive = Alive carp = Cardiovascular or pulmonary lunc = Lung cancer other = All other causes of death miss = Unknown cause of death

Variable name	*Meaning and response codes*
22) EDD	Less than high school education (0 = No, 1 = Yes)
23) EDC	Less than 4-yr college graduate (0 = No, 1 = Yes) NB: SAS label in mort6c.ssd is wrong
24) ED0	Did not complete grade school (0 = No, 1 = Yes)
25) ED1	Grade school completed and no further education (0 = No, 1 = Yes)
26) ED2	High school completed and no further education (0 = No, 1 = Yes)
27) ED3	Trade school or only attended college and no further education (0 = No, 1 = Yes)
28) ED4	2-yr college or nursing graduate and no further education (0 = No, 1 = Yes)
29) ED5	4-yr college graduate and no further education (0 = No, 1 = Yes)
30) ED6	Post-graduate (0 = No, 1 = Yes)
31) WAT	Watertown participant (0 = No, 1 = Yes)
32) KIN	Kingston / Harriman participant (0 = No, 1 = Yes)
33) STL	St. Louis participant (0 = No, 1 = Yes)
34) STE	Steubenville participant (0 = No, 1 = Yes)
35) POR	Portage participant (0 = No, 1 = Yes)
36) TOP	Topeka participant (0 = No, 1 = Yes)
37) BMI	Body mass index (kg/m**2)
38) FSMOKE	Former smoker but not current (0 = No, 1 = Yes)
39) CPACK	Current smoker pack years
40) FPACK	Former smoker pack years
41) SMOKEST	Smoking status cs = current smoker fs = former smoker ns = never smoked
42) DUST	Job exposure to dust (0 = No, 1 = Yes)
43) FUME	Job exposure to fumes (0 = No, 1 = Yes)
44) OCCU	Occupational exposure to dust or fumes (0 = No, 1 = Yes)
45) SO4	SO4/8 (ug/m3)
46) PM15	PM15/28.3 (ug/m3)
47) PM2_5	PM2_5/18.6 (ug/m3)
48) TSP	TSP/55.8 (ug/m3)
49) SO2	SO2/28.3 (ppb)
50) H	H+/25.8 (nmol/m3)
51) O3	Daily Mean Ozone (ppb)
52) O3M	Daily max ozone (ppb)
53) NO2	NO2 (ppb)/15.8 (range)
54) DUSTPM	Dust * PM2_5

APPENDIX D. Questionnaires and Codebook Used in the American Cancer Society Study

AMERICAN CANCER SOCIETY CANCER PREVENTION STUDY II QUESTIONNAIRE FOR MEN 	Division No.	Unit No.	Group No.
	Researcher No.	Family No.	Person No.

Date: _____

1. Name: _____
2. Date of birth: Month _____ Year _____
3. How old are you now? _____
4. Current weight with indoor clothing: _____ lbs.
5. Weight 1 year ago: _____ lbs.
6. Height (without shoes): _____ ft. _____ in.
7. ☐ White ☐ Black ☐ Hispanic ☐ Oriental ☐ Other _____ (specify)
8. Marital status: ☐ Single ☐ Separated ☐ Widowed ☐ Married ☐ Divorced
9. If ever married, age at first marriage: _____
10. Number of times married: _____
11. Social Security No.: _____ (optional)

FAMILY HISTORY (IN RELATION TO CANCER):

1. Fill in the following table as completely as possible for parents, brothers and sisters.

LIST ONE BLOOD RELATIVE PER LINE: (Circle Brother **or** Sister)	IS THIS PERSON? (Circle One)	IF ALIVE, GIVE AGE	IF DEAD, GIVE AGE AT DEATH	DID THIS PERSON **EVER** HAVE CANCER? (Circle One)	IF "YES," SPECIFY TYPE OF CANCER	AT WHAT AGE?
Father	Alive Dead			Yes No		
Mother	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		

2. When you were born, a) How old was your mother? _____ b) How old was your father? _____

HISTORY OF DISEASES:

1. Have you ever had cancer? ☐ Yes ☐ No. If "yes,"
   a) What type? _____
   b) Date of first treatment: _____
2. Place a check-mark by the following diseases or conditions for which you have ever been diagnosed by a doctor:
   - ☐ High Blood Pressure
   - ☐ Heart Disease
   - ☐ Stroke
   - ☐ Diabetes
   - ☐ Gall Stones
   - ☐ Chronic Indigestion
   - ☐ Kidney Disease
   - ☐ Kidney Stones
   - ☐ Bladder Disease
   - ☐ Cirrhosis of the Liver
   - ☐ Tuberculosis
   - ☐ Chronic Bronchitis
   - ☐ Emphysema
   - ☐ Hay Fever
   - ☐ Asthma
   - ☐ Stomach Ulcer
   - ☐ Duodenal Ulcer
   - ☐ Diverticulosis
   - ☐ Rectal Polyps
   - ☐ Colon Polyps
   - ☐ Thyroid Condition
   - ☐ Arthritis
   - ☐ Prostate Trouble
   - ☐ Hepatitis
   - ☐ Any other serious disease (specify) _____
3. Have you ever had an operation? ☐ Yes ☐ No
   If "yes," specify type and date(s) of operation(s):
   _____
   _____
   _____
4. How **many** x-ray or fluoroscopic **examinations** (GI series, barium enema, etc.) have you **ever** had of:

	0	1-5	6 or More		0	1-5	6 or More
Stomach	☐	☐	☐	Chest	☐	☐	☐
Intestine	☐	☐	☐	Arms/Legs	☐	☐	☐
Back	☐	☐	☐	Head/Neck	☐	☐	☐

5. Have you ever been **treated** with radium, x-rays, or radioactive isotopes? ☐ Yes ☐ No
   If "yes," when? _____
   For what disease? _____
   _____
   What part of your body? _____
   _____
6. How many times have you had colds or flu in the past twelve months? _____

CURRENT PHYSICAL CONDITION:

1. How much exercise do you get (work or play)?
   ☐ None ☐ Slight ☐ Moderate ☐ Heavy
2. On the average, how many hours do you sleep each night?_____
3. On the average, how many times a month do you have insomnia?_____ ☐ None
4. Within the last month, have you noticed:
   a) Painful or frequent urination? ☐ Yes ☐ No
   b) An unusual discharge from your penis?
   ☐ Yes ☐ No
5. Do you notice pains in your legs when you walk which go away when you rest? ☐ Yes ☐ No
   If "yes," how many years have you had these pains?_____
6. Are you sick at the present time? ☐ Yes ☐ No
   If "yes," with what disease or condition?_____
   _____
   _____
   _____

HABITS:

1. **Whether or not you smoke,** on the average, how many **hours a day** are you exposed to cigarette smoke of others:
   At home_____, At work_____, In other areas_____.
2. Do you now or have you ever smoked cigarettes, cigars or pipes, at least one a day for one year's time? ☐ Yes ☐ No
   If never smoked, skip to question 8.
3. If you **currently** smoke cigarettes, cigars or pipes, fill in the information below:

Current Smokers	Cigarettes	Cigars	Pipes
Average number smoked per day			
Age began smoking			
INHALATION:			
Do not inhale			
Inhale slightly			
Inhale moderately			
Inhale deeply			
Total years of smoking			
Years smoked **filtered** cigarettes			
Years smoked **non-filtered** cigarettes			

4. Current brand of cigarette:_____
   a) Size: ☐ Regular ☐ King ☐ 100 mm ☐ 120 mm
   b) ☐ Non-filter ☐ Filter ☐ Menthol
   c) Years smoked this brand:_____

5. If you have **quit** smoking cigarettes, cigars or pipes, fill in the information below:

Ex-Smokers	Cigarettes	Cigars	Pipes
Average number smoked per day			
Age began smoking			
Age quit			
INHALATION:			
Did not inhale			
Inhaled slightly			
Inhaled moderately			
Inhaled deeply			
Total years smoked			
Years smoked **filtered** cigarettes			
Years smoked **non-filtered** cigarettes			

6. Last brand of cigarette smoked:_____
   a) Size: ☐ Regular ☐ King ☐ 100 mm
   ☐ 120 mm
   b) ☐ Non-filter ☐ Filter ☐ Menthol
   c) Years smoked this brand:_____

7. Current **and** ex-cigarette smokers, fill in the following information for:
   1) The **first** brand smoked regularly; and
   2) The brand of cigarette smoked for the **longest** period of time.

Brand Name	Size	Filter		Menthol		Number Per Day	Years
		Yes	No	Yes	No		
1.							
2.							

8. Have you ever chewed tobacco at least once a week for at least one year? ☐ Yes ☐ No
   If "no," skip to question 9.
   a) Age began chewing tobacco:_____
   b) How many times a week?_____
   c) For how many years?_____
   d) Do you still chew tobacco? ☐ Yes ☐ No

9. Have you ever used snuff at least once a week for at least one year? ☐ Yes ☐ No
   If "no," skip to "Diet."
   a) Age began using snuff:_____
   b) How many times a week?_____
   c) For how many years?_____
   d) Do you still use snuff? ☐ Yes ☐ No

DIET:

1. On the average, how many days per week do you eat the following foods? (If less than once a week, but at least twice a month, write 1/2.)

Beef	Raw vegetables
Pork	Carrots
Chicken	Squash/Corn
Liver	Citrus fruits/Juices
Ham	Spaghetti/Macaroni/ White rice
Fish	White bread/Rolls/ Biscuits
Smoked meats	Brown rice/Whole wheat/Barley
Frankfurters/ Sausage	Bran/Corn muffins
Butter	Potatoes
Margarine	Oatmeal/Shredded wheat/Bran cereals
Cheese	Cold (Dry) cereals
Eggs	Ice cream
Green leafy vegetables	Chocolate
Tomatoes	
Cabbage/Broccoli/ Brussels sprouts	

2. How many days a week do you eat the following **fried** foods?

Fried eggs	Fried hamburgers or beef
Fried bacon	Other fried foods
Fried chicken/fish	
French fries	

DO NOT EAT FRIED FOODS ☐

3. Do you eat a vegetarian diet? ☐ Yes ☐ No
   If "yes," what type and for how many years? _____

4. Has there been a major change in your diet in the last 10 years? ☐ Yes ☐ No
   If "yes," what was the change? _____

5. a) Do you now or have you ever added artificial sweeteners (saccharin or cyclamates) to coffee, tea, or other drinks or food?
   ☐ Yes, currently ☐ Formerly ☐ Never
   b) If **ever** used artificial sweeteners, indicate amount per day and for how long.
   Packets: No. per day _____ Years _____
   Drops: No. per day _____ Years _____
   Tablets: No. per day _____ Years _____

6. Do you get your drinking water from: ☐ City supply ☐ Private well ☐ Other (specify) _____

7. Do you add any substances to soften your drinking water? ☐ Yes ☐ No

8. How many cups, glasses, or drinks of these beverages do you usually drink a day, and for how many years? (If you no longer drink a listed beverage, or your pattern has changed in the last ten years, indicate previous and current amounts. If less than once a day, but at least three times a week, write 1/2.)

Beverages	Currently		Previously	
	Amount	Years	Amount	Years
Whole milk (not skim milk)				
Caffeinated coffee				
Decaffeinated coffee				
Tea				
Diet soda or diet iced tea				
Non-diet colas				
Other non-diet soft drinks				
Beer				
Wine				
Hard liquor				

MEDICATIONS AND VITAMINS:

1. How many times in the last month have you used the following and how long have you used them? (If none, write 0; if used only occasionally, write 1/2.)

Medications and Vitamins	Times	Years
Aspirin, Bufferin, Anacin		
Tylenol		
Vitamin A		
Vitamin C		
Vitamin E		
Multi-Vitamins		
Blood Pressure pills		
Diuretics (water pills)		
Thyroid medications		
Heart medications		
Anti-Acid medications		
Valium		
Librium		
Prescription sleeping pills		
Tagamet (for ulcers)		
Other: _____		
_____		

OCCUPATIONS:

1. What is your current occupation and what are your duties? _____
   _____
   _____ How many years: _____
2. If retired, what was your last occupation? _____
   _____
   _____
   _____ Year retired: _____
3. What other job have you held for the longest period of time? _____
   _____
   _____ How many years: _____
4. What time of day do you start working? _____
   Do you work rotating shifts? ☐ Yes ☐ No
5. How many hours a week do you work on:
   paid jobs _____, volunteer work _____,
   housework _____.
6. In your work or daily life, are (were) you **regularly** exposed to any of the following? If "yes," indicate the number of years exposed.

Exposure to:	Check One		Number of Years
	Yes	No	
Asbestos			
Chemicals/Acids/Solvents			
Coal or Stone Dusts			
Coal Tar/Pitch/Asphalt			
Diesel Engine Exhaust			
Dyes			
Formaldehyde			
Gasoline Exhaust			
Pesticides/Herbicides			
Textile Fibers/Dusts			
Wood Dust			
X-rays/Radioactive Materials			

REMARKS:

MISCELLANEOUS:

1. Where were you born? _____ (city) _____ (state/country)
2. Where were your parents born?
   Father: _____
   Mother: _____
3. Religion: ☐ Protestant ☐ Catholic ☐ Jewish
   ☐ LDS ☐ Other _____ ☐ None
   If Protestant, what denomination? _____
4. Education:
   ☐ 8th Grade or Less ☐ Some College
   ☐ Some High School ☐ College Graduate
   ☐ High School Graduate ☐ Graduate School
   ☐ Vocational/Trade School
5. How many years have you lived in your present neighborhood? _____
6. How many friends or relatives do you feel close to? _____
7. How many times a month do you:
   a) Go to church or temple? _____
   b) Attend club meetings? _____
   c) Participate in group activities? _____
8. Were you in the U.S. Armed Services? ☐ Yes ☐ No
   If "yes,"
   a) What branch of the service were you in? _____
      _____
   b) What were your dates of service?
      _____ to _____,
      _____ to _____.
   c) Where did you serve? _____
      _____
9. What is the most upsetting event that happened to you in about the last five years? _____
   _____
   _____ ☐ None
10. Do you now or have you ever used mouthwash? ☐ Yes ☐ No
    If "yes,"
    a) What brand? _____
    b) How many times a week is it used? _____
    c) For how many years have you used it? _____

AMERICAN CANCER SOCIETY **CANCER PREVENTION STUDY II** **QUESTIONNAIRE FOR WOMEN** 	Division No.	Unit No.	Group No.
	Researcher No.	Family No.	Person No.

Date: _____

1. Name: _____
2. Date of birth: Month _____ Year _____
3. How old are you now? _____
4. Current weight with indoor clothing: _____ lbs.
5. Weight 1 year ago: _____ lbs.
6. Height (without shoes): _____ ft. _____ in.
7. ☐ White ☐ Black ☐ Hispanic ☐ Oriental ☐ Other _____ (specify)
8. Marital status: ☐ Single ☐ Separated ☐ Widowed ☐ Married ☐ Divorced
9. If ever married, age at first marriage: _____
10. Number of times married: _____
11. Social Security No.: _____ (optional)

FAMILY HISTORY (IN RELATION TO CANCER):

1. Fill in the following table as completely as possible for parents, brothers and sisters.

LIST ONE BLOOD RELATIVE PER LINE: (Circle Brother **or** Sister)	IS THIS PERSON? (Circle One)	IF ALIVE, GIVE AGE	IF DEAD, GIVE AGE AT DEATH	DID THIS PERSON **EVER** HAVE CANCER? (Circle One)	IF "YES," SPECIFY TYPE OF CANCER	AT WHAT AGE?
Father	Alive Dead			Yes No		
Mother	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		
Brother or Sister	Alive Dead			Yes No		

2. When you were born, a) How old was your mother? _____ b) How old was your father? _____

HISTORY OF DISEASES:

1. Have you ever had cancer? ☐ Yes ☐ No. If "yes,"
   a) What type? _____
   b) Date of first treatment: _____
2. Place a check-mark by the following diseases or conditions for which you have ever been diagnosed by a doctor:
   - ☐ High Blood Pressure
   - ☐ Heart Disease
   - ☐ Stroke
   - ☐ Diabetes
   - ☐ Gall Stones
   - ☐ Chronic Indigestion
   - ☐ Kidney Disease
   - ☐ Kidney Stones
   - ☐ Bladder Disease
   - ☐ Cirrhosis of the Liver
   - ☐ Tuberculosis
   - ☐ Chronic Bronchitis
   - ☐ Emphysema
   - ☐ Hay Fever
   - ☐ Asthma
   - ☐ Stomach Ulcer
   - ☐ Duodenal Ulcer
   - ☐ Diverticulosis
   - ☐ Rectal Polyps
   - ☐ Colon Polyps
   - ☐ Thyroid Condition
   - ☐ Arthritis
   - ☐ Breast Cysts
   - ☐ Gynecological Problems
   - ☐ Hepatitis
   - ☐ Any other serious disease (specify) _____
3. Have you ever had an operation? ☐ Yes ☐ No
   If "yes," specify type and date(s) of operation(s):
4. How **many** x-ray or fluoroscopic **examinations** (GI series, barium enema, etc.) have you **ever** had of:

	0	1-5	6 or More		0	1-5	6 or More
Stomach	☐	☐	☐	Chest	☐	☐	☐
Intestine	☐	☐	☐	Breast	☐	☐	☐
Back	☐	☐	☐	Head/Neck	☐	☐	☐

5. Have you ever been **treated** with radium, x-rays, or radioactive isotopes? ☐ Yes ☐ No
   If "yes," when? _____
   For what disease? _____
   What part of your body? _____
6. How many times have you had colds or flu in the past twelve months? _____

CURRENT PHYSICAL CONDITION:

1. How much exercise do you get (work or play)?
   ☐ None ☐ Slight ☐ Moderate ☐ Heavy
2. On the average, how many hours do you sleep each night?_____
3. On the average, how many times a month do you have insomnia?_____ ☐ None
4. Within the last twelve months, have you noticed:
   a) A lump or thickening in your breast?
   ☐ Yes ☐ No
   b) An unusual discharge from your breast?
   ☐ Yes ☐ No
5. Do you notice pains in your legs when you walk which go away when you rest? ☐ Yes ☐ No
   If "yes," how many years have you had these pains?_____
6. Are you sick at the present time? ☐ Yes ☐ No
   If "yes," with what disease or condition?_____
   _____
   _____

MENSTRUAL AND REPRODUCTIVE HISTORY:

1. How old were you when menstruation began?____
2. What is your current menopausal status?
   ☐ Still regularly menstruating
   ☐ In menopause ☐ Past menopause
3. During your menstrual history:
   a) Are (were) your periods: ☐ Regular ☐ Irregular
   b) What is (was) the usual number of days of flow?_____
4. **If past menopause:**
   a) Was your menopause: ☐ Natural ☐ Artificial
   b) Age when periods stopped completely?____
   c) Did you have excessive bleeding during menopause? ☐ Yes ☐ No
5. Have you ever had or tried to have children?
   ☐ Yes ☐ No
   If "no," skip to question 9.
6. Have you ever had difficulty becoming pregnant?
   ☐ Yes ☐ No
   If "yes," what was the reason?_____
   _____
   _____
7. How many times have you been pregnant?____
   a) Your age at your first pregnancy?_____
   b) Your age at your first live birth?_____
   c) Number of children born alive?_____
   d) Number of stillbirths (carried 5 months or more)?_____
   e) Number of miscarriages (carried less than 5 months)?_____
8. Were you ever given DES (Diethylstilbestrol) to prevent miscarriage? ☐ Yes ☐ No
   If "yes,"
   a) At what age did you take it?_____
   b) For how many months did you take it?_____
9. Birth control methods: Indicate your age when **first used** and number of years of use.

Method Used	Age	Years
Rhythm		
Diaphragm		
Cream/Foam/Jelly		
Tubal Ligation		
Intrauterine Device (IUD)		
Condom (partner)		
Vasectomy (partner)		
NONE OF THE ABOVE ☐		

10. Have you **ever** taken oral contraceptives (birth control pills)? ☐ Yes ☐ No
    If "no," skip to question 11.
    a) Age when you first took them?_____
    b) How many years did you take them?_____
    c) What brand(s) do (did) you take?_____
    _____
    d) If you stopped taking them, what was the reason?_____
    e) Did you have irregular or painful periods when you stopped? ☐ Yes ☐ No
11. Have you **ever** used female hormones (estrogens) other than oral contraceptives? ☐ Yes ☐ No
    a) Why do (did) you take estrogens?
    ☐ Menopausal symptoms ☐ Hysterectomy
    ☐ Bone problems ☐ Cancer
    ☐ Other (specify)_____
    b) Age first took estrogens?_____
    c) For how many years did you take them?_____
    d) How did you take them? ☐ Injection ☐ Cream
    ☐ Pill (brand):_____

HABITS:

1. **Whether or not you smoke,** on the average, how many **hours a day** are you exposed to cigarette smoke of others:
   At home____, At work____, In other areas____.
2. Do you now or have you ever smoked cigarettes, at least one a day for one year's time? ☐ Yes ☐ No

Smoking History	Current Smokers	Ex-Smokers
Number smoked a day		
Age began smoking		
Age quit smoking		
Most recent (last) brand		
Years smoked **this** brand		
Total years smoked **filtered** cigarettes		
Total years smoked **non-filtered** cigarettes		
Total years of smoking (filtered + non-filtered)		

3. Current **and** ex-smokers:
   a) Do (did) you inhale? ☐ No, never ☐ Slightly ☐ Moderately ☐ Deeply
   b) Fill in the following information for:
      1) The **first** brand smoked regularly; and
      2) The brand of cigarette smoked for the **longest** period of time.

Brand Name	Size	Filter		Menthol		Number Per Day	Years
		Yes	No	Yes	No		
1.							
2.							

DIET:

1. On the average, how many days per week do you eat the following foods? (If less than once a week, but at least twice a month, write 1/2.)

Beef		Raw vegetables	
Pork		Carrots	
Chicken		Squash/Corn	
Liver		Citrus fruits/Juices	
Ham		Spaghetti/Macaroni/ White rice	
Fish			
Smoked meats		White bread/Rolls/ Biscuits	
Frankfurters/ Sausage		Brown rice/Whole wheat/Barley	
Butter		Bran/Corn muffins	
Margarine		Potatoes	
Cheese		Oatmeal/Shredded wheat/Bran cereals	
Eggs			
Green leafy vegetables		Cold (Dry) cereals	
Tomatoes		Ice cream	
Cabbage/Broccoli/ Brussels sprouts		Chocolate	

2. How many days a week do you eat the following **fried** foods?

Fried eggs		Fried hamburgers or beef	
Fried bacon			
Fried chicken/fish		Other fried foods	
French fries			

   DO NOT EAT FRIED FOODS ☐

3. Do you eat a vegetarian diet? ☐ Yes ☐ No
   If "yes," what type and for how many years? _____

4. Has there been a major change in your diet in the last 10 years? ☐ Yes ☐ No
   If "yes," what was the change? _____

5. a) Do you now or have you ever added artificial sweeteners (saccharin or cyclamates) to coffee, tea, or other drinks or food?
   ☐ Yes, currently ☐ Formerly ☐ Never
   b) If **ever** used artificial sweeteners, indicate amount per day and for how long.
   Packets: No. per day _____ Years _____
   Drops: No. per day _____ Years _____
   Tablets: No. per day _____ Years _____

6. Do you get your drinking water from: ☐ City supply ☐ Private well ☐ Other (specify) _____

7. Do you add any substances to soften your drinking water? ☐ Yes ☐ No

8. How many cups, glasses, or drinks of these beverages do you usually drink a day, and for how many years? (If you no longer drink a listed beverage, or your pattern has changed in the last ten years, indicate previous and current amounts. If less than once a day, but at least three times a week, write 1/2).

Beverages	Currently		Previously	
	Amount	Years	Amount	Years
Whole milk (not skim milk)				
Caffeinated coffee				
Decaffeinated coffee				
Tea				
Diet soda or diet iced tea				
Non-diet colas				
Other non-diet soft drinks				
Beer				
Wine				
Hard liquor				

MEDICATIONS AND VITAMINS:

1. How many times in the last month have you used the following and how long have you used them? (If none, write 0; if used only occasionally, write 1/2.)

Medications and Vitamins	Times	Years
Aspirin, Bufferin, Anacin		
Tylenol		
Vitamin A		
Vitamin C		
Vitamin E		
Multi-Vitamins		
Blood Pressure pills		
Diuretics (water pills)		
Thyroid medications		
Heart medications		
Anti-Acid medications		
Valium		
Librium		
Prescription sleeping pills		
Tagamet (for ulcers)		
Other: _____		
_____		

OCCUPATIONS:

1. What is your current occupation and what are your duties? _____
   _____
   _____ How many years: _____
2. If retired, what was your last occupation? _____
   _____
   _____
   _____ Year retired: _____
3. What other job have you held for the longest period of time? _____
   _____
   _____ How many years: _____
4. What time of day do you start working? _____
   Do you work rotating shifts? ☐ Yes ☐ No
5. How many hours a week do you work on:
   paid jobs _____, volunteer work _____,
   housework _____.
6. In your work or daily life, are (were) you **regularly** exposed to any of the following? If "yes," indicate the number of years exposed.

Exposure to:	Check One		Number of Years
	Yes	No	
Asbestos			
Chemicals/Acids/Solvents			
Coal or Stone Dusts			
Coal Tar/Pitch/Asphalt			
Diesel Engine Exhaust			
Dyes			
Formaldehyde			
Gasoline Exhaust			
Pesticides/Herbicides			
Textile Fibers/Dusts			
Wood Dust			
X-rays/Radioactive Materials			

REMARKS:

MISCELLANEOUS:

1. Where were you born? _____ (city) _____ (state/country)
2. Where were your parents born?
   Father: _____
   Mother: _____
3. Religion: ☐ Protestant ☐ Catholic ☐ Jewish
   ☐ LDS ☐ Other _____ ☐ None
   If Protestant, what denomination? _____
4. Education:
   ☐ 8th Grade or Less ☐ Some College
   ☐ Some High School ☐ College Graduate
   ☐ High School Graduate ☐ Graduate School
   ☐ Vocational/Trade School
5. How many years have you lived in your present neighborhood? _____
6. How many friends or relatives do you feel close to? _____
7. How many times a month do you:
   a) Go to church or temple? _____
   b) Attend club meetings? _____
   c) Participate in group activities? _____
8. What is the most upsetting event that happened to you in about the last five years? _____
   _____
   _____ ☐ None
9. How many people do you take care of in your household? (Include yourself) _____
10. Do you now or have you ever used a **permanent** hair dye? ☐ Yes ☐ No
    If "yes,"
    a) What brand? _____
    b) What color? _____
    c) How often applied? _____
    d) How many years have you used it? _____
11. Do you now or have you ever used mouthwash? ☐ Yes ☐ No
    If "yes,"
    a) What brand? _____
    b) How many times a week is it used? _____
    c) For how many years have you used it? _____

ACSS Re-analysis Codebook for RAWDATA

ACSS Re-analysis Codebook for RAWDATA

(from CPSII, Poll., Mort. Fail. And Climate)

Observations: 684,296

Variable Name	Meaning and response codes
Data from CPSII (Observations: 684,296)	
1) ID	14-digit, CPS-II ID
2) CANSITE	Site of Cancer (00 - 99, N, U)
3) AGE_INT	Age at Interview
4) RACE	
	1 = White
	2 = White and Hispanic
	3 = Black
	4 = Black and Hispanic
	5 = Hispanic
	6 = Asian
	7 = Other
5) SEX	1 = Male
	2 = Female
6) EDUCATE	1 = 8th grade or less
	2 = Some high school
	3 = High school graduate
	4 = Vocational/trade school
	5 = Some college
	6 = College graduate
	7 = Graduate school
7) BMI	Body Mass Index (Kg/M^2)
8) ASBESTOS	1 = Exposed to asbestos
	2 = No.
9) CHEMICAL	1 = Exposed to chemicals, acids, solvents
	2 = No.
10) COALDUST	1 = Exposed to coal/stone dust
	2 = No.
11) COALTAR	1 = Exposed to coal tar, pitch, asphalt
	2 = No.
12) DIESEL	1 = Exposed to diesel engine exhaust
	2 = No.
13) FORHYDE	1 = Exposed to formaldehyde
	2 = No.
14) BEERC	Beer, current amount
15) LIQC	Hard liquor, current amount

ACSS Re-analysis Codebook for RAWDATA

16) WINEC	Wine, current amount
17) PSMKHM	Passive smoking at home (hours per day)
18) PSMKWK	Passive smoking at work (hours per day)
19) PSMKOTH	Passive smoking elsewhere (hours per day)
20) SMKSTAT	1 = Never smoker
	2 = Current cigarette
	3 = Current cig/pipe smoker(Male), Ex-smoker(Female)
	4 = Pipe/cigar smoker only
	5 = Ex-cigarette smoker
	6 = Ex-cig/pipe smoker
	7 = Ex-pipe/cigar smoker only
	8 = Ex-cig smoker, current pipe/cigar smoker
21) SMKCAGE	Age of starting smoking for current cigarette smokers
22) SMKCPD	Number of current cigarette smokers per day
23) SMKCYR	Years of current cigarette smoking
24) SMKCQUIT	Age of quitting smoking for current cigarette smokers
25) XSMKCAGE	Age of starting smoking for former cigarette smokers
26) XSMKCPD	Number of former cigarette smokers per day
27) XSMKCYR	Years of former cigarette smoking
28) FLAGDEL	0 = No missing data on factors evaluated
	1 = Missing in race
	2 = Missing in smoker or in ex-smoker
	3 = Missing in educate
	4 = Missing in BMI
	5 = Missing in DIV
	6 = Missing in Passive smoking
	7 = Missing in others
29) VS	D = Dead, reported in 1984
	G = Dead, reported in 1986
	K = Dead, reported in 1988
	N = Dead, former lost-to-follow-up 1982 - 1988
	L = Dead, NDI follow-up since 1988
	P = Dead, pending diagnosis
	. = Alive, or lost-to follow-up

Data from MORT

30) CODE1	Code of death (ICD9 - codes)
31) CODETYPE	Death code indicator
	1 = Combined CODE1 with two codes position
	2 = Combined CODE1 with four codes position
32) DEATH_MO	Month of death
33) DEATH_YR	Year of death

ACSS Re-analysis Codebook for RAWDATA

Data from FAIL

34) FAIL	Time on the follow-up in months

Data from POLL

35) FP	Mean fine particulate
36) FPF	Median fine particulate
37) FPFDEL	0 = No FPF missing
	1 = FPF missing
38) MEANSULF	Mean sulfates
39) SULFDEL	0 = No MEANSULF missing
	1 = MEANSULF missing
40) LOGDEN	
41) POVERTY	
42) SO4AST	
43) SO4SA	
44) ST	2-char, States name
45) NAME	4-char, Areas name

Climate Data (observations = 157)

46) TEMPMEAN	Mean temperature
47) TEMPMIN	Minimum temperature
48) TEMPMAX	Maximum temperature
49) DCOLD	Dummy coding for mean temperature
	0 = Greater or equal 50.0^0 F
	1 = Less than 50.0^0 F
50) DHOT	Indicator TEMPMEAN
	0 = Equal or less than 60.0^0 F
	1 = Greater than 60.0^0 F

PART I APPENDICES AVAILABLE ON REQUEST

The following Appendices may be obtained by contacting the Health Effects Institute at 955 Massachusetts Avenue, Cambridge MA 02139, by phone (617-876-6700), fax (617-876-6709), or e-mail (pubs@healtheffects.org). Please give the full title of the Special Report, the Part I title, and the titles of the Appendices you wish to request.

E. Computer Programs and Output Used in the Replication of the Original Analyses of the Harvard Six Cities Study

F. Computer Programs and Output Used in the Replication of the Original Analyses of the American Cancer Society Study

Acknowledgments

Daniel Krewski, Richard T Burnett, Mark S Goldberg, Kristin Hoover, Jack Siemiatycki,
Michael Jarrett, Michal Abrahamowicz, Warren H White, and Others

This reanalysis of the Six Cities and ACS Studies of the association between particulate air pollution and mortality was a complex undertaking, involving a large number of scientists representing a range of disciplines. The Reanalysis Team itself comprised 31 individuals from 13 institutions in Canada and the United States; the new Centre for Population Health Risk Assessment in the Institute of Population Health at the University of Ottawa served as the focal point for the project.

All members of the Reanalysis Team made unique contributions to this multidisciplinary effort and are acknowledged elsewhere by name. Although all team members made outstanding contributions, Alette Willis deserves special recognition not only for her painstaking technical work in reconstructing the 1980 boundaries of the metropolitan areas included in the ACS Study and assembling additional socioeconomic, demographic, and environmental data for use by the Reanalysis Team, but also for serving as technical editor and coordinator for various report drafts. Kate Keating and Paula Carty also assisted with technical editing of report drafts, all of which were prepared under tight timelines. Additional contributions were made by Kimberly Zartolas and Ashley Wong, who spent the summer of 1999 at Harvard University coding the residence histories for all subjects included in the Six Cities Study, and by Sylvie Mauviel, who provided secretarial and administrative services to team members.

The Reanalysis Team gratefully acknowledges the cooperation of the Original Investigators in both the Six Cities Study and the ACS Study. Dr Douglas W Dockery, Dr Frank E Speizer, and Martha Fay answered many questions from the Reanalysis Team about the Six Cities Study, as did Drs C Arden Pope III, Michael J Thun, and Eugenia Calle about the ACS Study. Site visits with both sets of Original Investigators early in the Reanalysis Project were also of great value. We could not have asked for more patience and cooperation throughout the two-year period during which the reanalysis was conducted.

The Reanalysis Team met with the HEI Expert Panel four times during the course of the project to review interim results and to discuss methodologic approaches to specific issues. Discussions with the Expert Panel were extremely valuable and served to clarify a number of critical analytic issues. The Reanalysis Team also benefitted greatly from formal comments provided by the Special Panel of the HEI Health Review Committee on drafts of our reports submitted to HEI. Although the Advisory Board did not have the same opportunities to review the work in progress, their comments on major analytic issues were of great value to the Reanalysis Team.

The important role of HEI staff members also needs to be acknowledged. Dr Aaron J Cohen was the research project manager at HEI and skillfully facilitated interactions among the Reanalysis Team, Expert Panel, and Advisory Board. We were pleased to work with such an experienced and knowledgeable research manager. Daniel S Greenbaum, President of HEI, made the Reanalysis Project a personal priority and was actively engaged throughout the process. Dr JoAnn Ten Brinke was the review project manager and oversaw the production of the final report at HEI, coordinating the efforts of the HEI Review Panel and the editorial staff in the final stages of the project. This was a complex undertaking, given the intensive level of review to which the draft and revised reports were subjected as well as the length and scope of the final report. The editorial staff at HEI, led by Virgi Hepner for this project, did an outstanding job of editing and preparing the final manuscript for publication.

A project of this magnitude could not have been completed in a timely manner without excellent administrative support. Howard E Garsh at HEI and Gilles Morier at the University of Ottawa handled the detailed administrative arrangements, permitting the Reanalysis Team to focus on the task at hand, knowing that budgets and schedules were being carefully monitored. Mariella Peca and Hélène l'Abbée worked closely with Gilles on administrative and financial project matters at the University of Ottawa.

A number of individuals less directly involved with the Reanalysis Project also need to be acknowledged. Bob Bilgrad at the National Death Index provided guidance on procedures to be followed in securing the approvals needed to work with data on the vital status of individual subjects. Drs John Bachman and John Vandenberg directed questions from the Reanalysis Team about monitoring data collected by the US Environmental Protection Agency to

the right scientists at EPA. We specifically acknowledge Jacob G Summers, who provided the Reanalysis Team with data from the Aerometric Information Retrieval System, and Jose M Sune, who provided us with data from the Inhalable Particle Monitoring Network.

About the Primary Authors

Daniel Krewski is currently professor of medicine and professor of epidemiology and community medicine at the University of Ottawa, where he is involved in a number of activities in population health risk assessment within the new Insitute of Population Health. Dr Krewski has also served as adjunt research professor of statistics in the Department of Mathematics and Statistics at Carleton University since 1984. Prior to joining the Faculty of Medicine at the University of Ottawa in 1998, he was director of Risk Management in the Health Protection Branch of Health Canada. While with Health Canada, he also served as acting director of the Bureau of Chemical Hazards and as chief of the Biostatistics Division in the Environmental Health Directorate. Dr Krewski obtained his PhD in statistics from Carleton University and subsequently completed an MHA at the University of Ottawa. His professional interests include epidemiology, biostatistics, risk assessment, and risk management. Dr Krewski is a Fellow of the American Statistical Association and the Society for Risk Analysis. He currently serves as a member of the US National Academy of Sciences' Board on Environmental Studies and Toxicology and as a member of the Academy's Committee on Research Priorities for Airborne Particulate Matter.

Richard T Burnett received his PhD from Queen's University in 1982 in mathematical statistics. He is now a senior research scientist with the Environmental Health Directorate of Health Canada, where he has been working since 1983 on issues relating to the health effects of outdoor air pollution. Dr Burnett's work has focused on the use of administrative health and environmental information to determine the public health impacts of combustion-related pollution using nonlinear random-effects models, time series and spatial analysis techniques.

Mark S Goldberg is an associate professor with the National Institute of Scientific Research at the University of Quebec in Laval and is adjunct professor at McGill University in Montreal. Dr Goldberg obtained degrees in physics (BS 1975) and epidemiology and biostatistics (MS 1985, PhD 1992) from McGill. He is currently investigating the short-term effects of air pollution on mortality and identifying susceptible subgroups. His research interests include occupational risk factors for female breast cancer, and delivery of treatment and other health services for breast cancer.

Kristin Hoover started Hoover Consultants in 1989 after a request for quality assurance consulting on the Natural Resources Damage Assessment for the *Valdez* oil spill. During the past 11 years, she has worked on many audits in toxicology, analytic chemistry, regulatory review, and quality assurance, especially the potential health effects of particulate and gaseous air pollution. Prior to consulting, Ms Hoover was corporate product safety manager for Pennwalt Corporation and a senior toxicologist at Arco Chemical Company. She holds degrees in natural resources and environmental biology and is senior author of the book Managing Conduct and Data Quality of Toxicology Studies, which was the culmination of a conference sponsored jointly by government and industry. Ms Hoover was a founding board member of the Society of Quality Assurance, which now counts over a thousand members.

Jack Siemiatycki is professor of epidemiology at the University of Quebec with a concurrent appointment at McGill University. His 150 publications and reports cover many methodologic and substantive issues, including health survey methodology, juvenile diabetes epidemiology, health effects of asbestos, and occupational causes of cancer. He is recognezed for the development of novel methods of exposure assessment for investigation of environmental/occupational causes of cancer. Dr Siemiatycki is a Distinguished Scientist of the Medical Research Council of Canada.

Michael Jerrett is assistant professor in the School of Geography and Geology, the Health Studies Program, and the Environmental Health Program at McMaster University, Hamilton, Ontario, Canada. He earned his PhD in geography from the University of Toronto. He specializes in spatial analysis of environment–health relationships. Dr Jerrett has designed and analyzed local, provincial and national level health and environment databases in Canada, the United States, Mexico, and Vietnam.

Michal Abrahamowicz is associate professor of biostatistics in the Department of Epidemiology and Biostatistics at McGill University and medical scientist in the Division of Clinical Epidemiology at Montreal General Hospital. His research interests include survival analyses, nonparametric regression, epidemiology of cancer, cardiovascular and rheumatologic diseases, cohort studies, and randomized clinical trials.

Warren H White is a senior research associate in the Chemistry Department at Washington University in St Louis. Educated as a mathematician at the California Institute of Technology and University of Wisconsin, Dr White spent several years teaching and researching in the area that has since become known as *chaos theory*. He returned to California Institute of Technology to work on the 1972–73 Aerosol Characterization Experiment (ACHEX), an early multiinvestigator field study of particulate air pollution in Los Angeles. Dr White has since continued to study ambient aerosols, their effects on radiative transfer and visibility, and their relationship to particle and precursor gas emissions. Dr White is a member of the EPA's Clean Air Science Advisory Committee and serves on National Academy of Science panels assessing research priorities for airborne particulate matter and tropospheric ozone.

Abbreviations and Other Terms

MEASURES OF PARTICLES AND SULFATE

$PM_{2.5}$	particulate matter 2.5 μm or smaller in aerodynamic diameter
$PM_{2.5}$(DC)	**mean** fine particle fraction from dichotomous samplers
$PM_{2.5}$(DC MD)	**median** fine particle mass concentration from dichotomous samplers
$PM_{2.5}$(OI MD)	**median** fine particle concentration used by the Original Investigators
PM_{10}	particulate matter 10 μm or smaller in aerodynamic diameter
PM_{15}	particulate matter 15 μm or smaller in aerodynamic diameter
PM_{15}(DC)	**mean** inhalable particle fraction from dichotomous samplers
PM_{15}(SSI)	**mean** inhalable particle fraction from high-volume SSI samplers
$PM_{15-2.5}$	the coarse particle fraction of particulate matter [15-μm particles minus 2.5-μm particles]
$PM_{15-2.5}$(DC)	**mean** coarse particle fraction from dichotomous samplers
SO_4^{2-}	sulfate
$SO_4^{2-}_{(\text{cb-adj region})}$	sulfate data for 1980–1981 inclusive, with region-specific adjustment for artifactual sulfate
$SO_4^{2-}_{(\text{cb-adj season})}$	sulfate data for 1980–1981 inclusive, with season-specific adjustment for artifactual sulfate
$SO_4^{2-}_{(\text{cb-adj US})}$	sulfate data for 1980–1981 inclusive, with US-specific adjustment for artifactual sulfate
$SO_4^{2-}_{(\text{cb-unadj})}$	sulfate data for 1980–1981 inclusive, unadjusted for artifactual sulfate
SO_4^{2-}(DC)	sulfate data from PM_{15}(DC)
SO_4^{2-}(OI)	sulfate data used by the Original Investigators
TSP	total suspended particles
TSP(IPMN)	**mean** TSP mass concentrations based on IPMN data

OTHER TERMS

ACS Study	the American Cancer Society Study
AIRS	Aerometric Information Retrieval System
ARRCCM	*American Review of Respiratory and Critical Care Medicine*
BMI	body mass index
$CaCO_3$	calcium carbonate
CAPITA	Center for Air Pollution Impact and Trend Analysis
CASAC	Clean Air Science Advisory Committee
CI	confidence interval
CO	carbon monoxide
CPS-II	American Cancer Society's Cancer Prevention Study II
DC	measurement from a dichotomous sampler
df	degrees of freedom
EPA	US Environmental Protection Agency
FP+CP	fine particles + coarse particles
FVC	forced vital capacity
H^+	aerosol acidity
HSPH	Harvard School of Public Health
IARC	International Agency for Research on Cancer
ICD-9	*International Classification of Diseases, Ninth Revision*
IP	inhalable particles
IPMN	Inhalable Particle Monitoring Network
JAWMA	*Journal of the Air and Waste Management Association*
MA	metropolitan area
MD	median
MSA	metropolitan statistical area
NAAQS	National Ambient Air Quality Standard
NAD	National Aerometric Database
NDI	National Death Index
NEJM	*New England Journal of Medicine*
NO_2	nitrogen dioxide

NOAA	US National Oceanic and Atmospheric Administration
O_3	ozone
OSI	Office of Scientific Integrity
r	bivariate correlation coefficient
range	the difference in mean concentrations between the most-polluted city and the least-polluted city
RR	relative risk
SAS	Statistical Application Software
SID	subject identification number
Six Cities Study	the Harvard Six Cities Study
SO_2	sulfur dioxide
SSI	high-volume sampler with size-selective inlet
SSN	Social Security Number

RELATED HEI PUBLICATIONS

PARTICULATE MATTER

Report Number	Title	Principal Investigator	Publication Date*
Research Reports			
75	Ozone Exposure and Daily Mortality in Mexico City: A Time-Series Analysis	DP Loomis	1996
81	Methods Development for Epidemiologic Investigations of the Health Effects of Prolonged Ozone Exposure	IB Tager	1998
82	Acute Effects of Ambient Ozone on Asthmatic, Wheezy, and Healthy Children	JM Peters	1998
83	Daily Changes in Oxygen Saturation and Pulse Rate Associated with Particulate Air Pollution and Barometric Pressure	DW Dockery	1999
86	Statistical Methods for Epidemiologic Studies of the Health Effects of Air Pollution	W Navidi	1999
91	Mechanisms of Morbidity and Mortality from Exposure to Ambient Air Particles	JJ Godleski	2000
93	Effects of Concentrated Ambient Particles in Rats and Hamsters: An Exploratory Study	T Gordon	2000
94	The National Morbidity, Mortality, and Air Pollution Study *Part I.* Methods and Methodologic Issues *Part II.* Morbidity, Mortality, and Air Pollution in the United States	JM Samet	2000
95	Association of Particulate Matter Components with Daily Mortality and Morbidity in Urban Populations	M Lippmann	2000
96	Acute Pulmonary Effects of Ultrafine Particles in Rats and Mice	G Oberdörster	2000
…	Daily Mortality and Fine and Ultrafine Particles in Erfurt, Germany *Part I.* Role of Particle Number and Particle Mass (in Press)	H-E Wichmann	2000
…	Identifying Subgroups of the General Population that May Be Susceptible to Short-Term Increases in Particulate Air Pollution: A Time-Series Study in Montreal, Quebec (in Press)	MS Goldberg	2000
…	A Case-Crossover Analysis of Fine Particulate Matter Air Pollution and Out-of-Hospital Sudden Cardiac Arrest (in Press)	H Checkoway	2000
Special Reports			
	Diesel Exhaust: A Critical Analysis of Emissions, Exposure, and Health Effects		1995
	Particulate Air Pollution and Daily Mortality: The Phase I Report of the Particle Epidemiology Evaluation Project		
	Phase I.A. Replication and Validation of Selected Studies		1995
	Phase I.B. Analyses of the Effects of Weather and Multiple Air Pollutants		1997
HEI Communications			
5	Formation and Characterization of Particles: Report of the 1996 HEI Workshop		1997
8	The Health Effects of Fine Particles: Key Questions and the 2003 Review		1999
HEI Program Summaries			
	Research on Particulate Matter		1999

** Reports published since 1990.*
Copies of these reports can be obtained by writing or calling the Health Effects Institute, 955 Massachusetts Avenue, Cambridge MA 02139. Phone (617) 876-6700 FAX (617) 876-6709 E-mail pubs@healtheffects.org www.healtheffects.org

BOARD OF DIRECTORS

Archibald Cox *Chair*
Carl M Loeb University Professor (Emeritus), Harvard Law School

Donald Kennedy *Vice Chair*
Editor-in-Chief, *Science*; President (Emeritus) and Bing Professor of Biological Sciences, Stanford University

Douglas Costle
Chairman of the Board and Distinguished Senior Fellow, Institute for Sustainable Communities

Alice Huang
Senior Councilor for External Relations, California Institute of Technology

Susan B King
Fellow, Sanford Institute of Public Policy, Duke University

Richard B Stewart
Professor, New York University School of Law

Robert M White
President (Emeritus), National Academy of Engineering, and Senior Fellow, University Corporation for Atmospheric Research

HEALTH RESEARCH COMMITTEE

Bernard D Goldstein *Chair*
Director, Environmental and Occupational Health Sciences Institute

Glen R Cass
Professor and Chairman, School of Earth and Atmospheric Sciences, Georgia Institute of Technology

Peter B Farmer
Professor and Section Head, Medical Research Council Toxicology Unit, University of Leicester

Seymour J Garte
Professor of Environmental and Community Medicine, Environmental and Occupational Health Sciences Institute

Helmut Greim
Professor and Chairman of Toxicology, Technical University Munich and GSF–National Research Center for Environment and Health

Rogene Henderson
Senior Scientist, Lovelace Respiratory Research Institute

Stephen I Rennard
Larson Professor, Pulmonary and Critical Care Medicine Section, Department of Internal Medicine, University of Nebraska Medical Center

Jonathan M Samet
Professor and Chairman, Department of Epidemiology, School of Public Health, Johns Hopkins University

Robert F Sawyer
Class of 1935 Professor of Energy (Emeritus), Professor of the Graduate School, University of California, Berkeley

Frank E Speizer
Edward H Kass Professor of Medicine, Channing Laboratory, Harvard Medical School and Department of Medicine, Brigham and Women's Hospital

Gerald van Belle
Professor, Departments of Environmental Health and Statistics, University of Washington

HEALTH REVIEW COMMITTEE

Daniel C Tosteson *Chair*
Professor of Cell Biology, Dean Emeritus, Harvard Medical School

Ross Anderson
Professor and Head, Department of Public Health Sciences, St George's Hospital Medical School, London University

John C Bailar III
Professor, Department of Health Studies, Biological Sciences Division, and Harris School of Public Policy, The University of Chicago

Thomas W Kensler
Professor, Division of Toxicological Sciences, Department of Environmental Sciences, Johns Hopkins University

Brian Leaderer
Professor, Department of Epidemiology and Public Health, Yale University School of Medicine

Thomas A Louis
Professor of Biostatistics, School of Public Health, University of Minnesota

Edo D Pellizzari
Vice President for Analytical and Chemical Sciences, Research Triangle Institute

Donald J Reed
Distinguished Professor of Biochemistry, Department of Biochemistry and Biophysics, and Environmental Health Sciences Center, Oregon State University

David J Riley
Professor of Medicine, University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School

Sverre Vedal
Professor of Medicine, University of British Columbia

OFFICERS & STAFF

Daniel S Greenbaum *President*
Robert M O'Keefe *Vice President*
Jane Warren *Director of Science*
Howard E Garsh *Director of Finance and Administration*
Richard M Cooper *Corporate Secretary*
Aaron J Cohen *Senior Scientist*
Maria G Costantini *Senior Scientist*
Bernard Jacobson *Staff Scientist*
Debra A Kaden *Senior Scientist*
Diane J Mundt *Staff Scientist*
Martha E Richmond *Staff Scientist*
Geoffrey H Sunshine *Staff Scientist*

JoAnn Ten Brinke *Staff Scientist*
Annemoon van Erp *Staff Scientist*
Gail V Allosso *Office and Contracts Manager*
Julia F Campeti *Scientific Copy Editor*
Sally Edwards *Managing Editor*
Terésa Fasulo *Senior Administrative Assistant*
L Virgi Hepner *Senior Scientific Editor*
Francine Marmenout *Senior Executive Assistant*
Teresina McGuire *Accounting Assistant*
Beverly Morse *Administrative Assistant*
Jacqueline C Rutledge *Controller*

HEALTH
EFFECTS
INSTITUTE

955 Massachusetts Avenue
Cambridge MA 02139 USA
+1-617-876-6700

SPECIAL REPORT

July 2000